

AD-A204 251

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

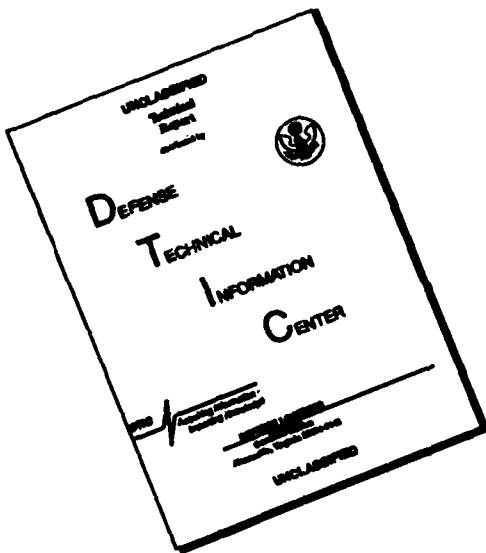
REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION <u>Unclassified</u>		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) JPL D-6027		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Jet Propulsion Laboratory California Institute of Technology		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) 4800 Oak Grove Drive Pasadena, CA 91109		7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MIPs 163-84, 142-85, 133-86	
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS		
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
			WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) REAL-TIME OPTICAL IMAGE PROCESSING TECHNIQUES			
12. PERSONAL AUTHOR(S) Liu, Hua-Kuang			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 84 July 24 TO 88 Oct 31	14. DATE OF REPORT (Year Month Day) 1988 October 31	15. PAGE COUNT 386
16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Nonlinear real-time optical image processing; optical processing; spatial light modulators; halftone screens; high-gamma spatial light modulators; liquid crystal television; spatial light modulators; pseudocolor image encoding; optical associative retrieval; neural processing. (LCLV)	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>Nonlinear real-time optical processing on spatial pulse frequency modulation has been pursued through the analysis, design, and fabrication of pulse frequency modulated halftone screens and the modification of micro-channel spatial light modulators (MSLMs). Micro-channel spatial light modulators are modified via the Fabry-Perot method to achieve the high gamma operation required for non-linear operation. Real-time nonlinear processing was performed using the halftone screen and MSLM. The experiments showed the effectiveness of the thresholding and also showed the needs of higher SBP for image processing.</p> <p>The Hughes LCLV has been characterized and found to yield high gamma (about 1.7) when operated in low frequency and low bias mode. Cascading of two LCLVs should also provide enough gamma for nonlinear processing. In this case, the SBP of the LCLV is sufficient but the uniformity of the LCLV needs improvement.</p> <p>The LCTV SLM has been found useful in a number of interesting real-time optical processing applications. These include image correlation, computer generation of holograms, pseudo-color image encoding for image enhancement, and associative-retrieval in neural processing.</p> <p>The discovery of the only known optical method for dynamic range compression of an input image in real-time by using GaAs photorefractive crystals is reported.</p> <p>Finally, a new architecture for non-linear multiple sensory, neural processing has been suggested for future work.</p>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

DISCLAIMER NOTICE



**THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE COPY
FURNISHED TO DTIC CONTAINED
A SIGNIFICANT NUMBER OF
PAGES WHICH DO NOT
REPRODUCE LEGIBLY.**

JPL D-6027

(2)

REAL-TIME OPTICAL IMAGE PROCESSING TECHNIQUES

FINAL REPORT OF THE TECHNICAL WORK

HUA-KUANG LIU

31 OCTOBER 1988

U. S. ARMY RESEARCH OFFICE

**ARO PROPOSAL NO. 22329-PH
MIPRs 163-84, 142-85 and 133-86**

**JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
4800 OAK GROVE DRIVE
PASADENA, CA 91109**

**APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.**

**DTIC
ELECTE
FEB 16 1989
S D
H**

Prepared by the Jet Propulsion Laboratory, California Institute of Technology,
for the U.S. Army Research Office through an agreement with the National
Aeronautics and Space Administration.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof.

This publication reports on work performed under NASA Contract No. NAS7-918, Task RE-182 Amendment No. 381, JPL Proposal No. 81-2405 and sponsored through ARO MIPRs 163-84, 142-85 and 133-86.

ABSTRACT

Nonlinear real-time optical processing on spatial pulse frequency modulation has been pursued through the analysis, design, and fabrication of pulse frequency modulated halftone screens and the modification of micro-channel spatial light modulators (MSLMs). Micro-channel spatial light modulators are modified via the Fabry-Perot method to achieve the high gamma operation required for non-linear operation. Real-time nonlinear processing was performed using the halftone screen and MSLM. The experiments showed the effectiveness of MSLM in its thresholding operation and also showed the needs of higher space bandwidth product (SBP) of the MSLM for image processing.

The Hughes LCLV has been characterized and found to yield high gamma (about 1.7) when operated in low frequency and low bias mode. Cascading of two LCLVs should also provide enough gamma for nonlinear processing. In this case, the SBP of the LCLV is sufficient but the uniformity of the LCLV needs improvement.

The LCTV SLM has been found useful in a number of interesting real-time optical processing applications. These include image correlation, computer-generation of holograms, pseudo-color image encoding for image enhancement, and associative-retrieval in neural processing.

The discovery of the only known optical method for dynamic range compression of an input image in real-time by using GaAs photorefractive crystals is reported.

Finally, a new architecture for non-linear multiple sensory, neural processing has been suggested for future work.



v

DTIC TAB	
Unannounced <input type="checkbox"/>	
Justification _____	
By _____	
Distribution/ _____	
Availability Codes _____	
Dist	Avail and/or Special
A-1	

ACKNOWLEDGEMENTS

The work on real-time optical processing described in this report was inspired, and financially supported by Bob Guenther, Physics Program Manager of the U.S. Army Research Office in Durham, North Carolina. His continued encouragement has been indispensable in helping the participants of this contract conduct research on the various real-time optical processing techniques.

The author would also like to thank the JPL management and staff members who have helped in producing this report. In particular, thanks are due to Y. Nakamura, H. Otake, A. Johnston, and P. LaFond for managing the contract and making valuable comments and suggestions; J. Brandt and M. F. Buehler for editing; and J. L. Donato and N. Parsons for typing of this report.

Finally, the author would like to thank the participants listed below for their fine contributions, and D. Z. Zhao for his help in the preparing of the proposal of this contract.

**PARTICIPANTS
AND
THEIR WORK ON THIS CONTRACT**

H. K. Liu	JPL Senior Research Engineer Principal Investigator
C. Warde	Professor, MIT (co-PI)
J. Davis	Professor, San Diego State University AO processing and LCLV SLM characterization
R. Lilly	Professor, San Diego State University AO processing and SLM characterization
S. Y. Kung	Professor, University of Southern Calif. New architecture for real-time processing
H. Chen	Professor, University of Alabama Screen simulation analysis
D. Chen	Professor, University of Alabama Screen simulation analysis
T. H. Chao	Member of Technical Staff, JPL Real-time processing
C. K. Chiang	Graduate Student, Univ. of Alabama Screen design and fabrication
H. Lamela-Rivera	Graduate Student, MIT MSLM
S. D. Law	Graduate Student, MIT MSLM and real-time processing
J. Diep	Optics Technician, JPL Experiment assistance

PREFACE

Real-time optical image processing is an important technique for military and on-board space applications such as the guidance and homing of missiles, and vision systems for space station robotics. Compact, lightweight, and high speed image processors are useful for on-board military surveillance, natural resources recognition, and space data-reduction and analysis.

For this purpose, a research project on real-time optical image processing techniques, supported by the Army Research Office in Durham, N.C., was performed at the Jet Propulsion Laboratory.

This final technical report is based on the work performed during the contract period. A total of around 10 publications have been credited fully or partially (as co-sponsor) to this ARO contract. The areas of the publications embraced halftone screen design, analysis, and fabrication; spatial light modulator research and development; real-time optical processing experiments; and new architectures for neural-net processing. A total of around 10 researchers from JPL and academia, including the University of Southern California, the University of Alabama, California State University at San Diego, and the Massachusetts Institute of Technology, have been partially supported by this project. One Ph.D thesis and one Master's thesis have resulted due to the partial support of this project.

The final report is organized in a way that essential results are properly integrated and placed in proper perspective with reports, reprints, and preprints of publications appended for easy reference.

TABLE OF CONTENTS

	<u>Page</u>
Report Documentation Page	i
Title Page	iii
Abstract	v
Acknowledgements	vi
Participants	vii
Preface	viii
Table of Contents	ix
Executive Summary	x
I. Introduction	1
II. Halftone Techniques for Spatial Pulse-Frequency Modulation	2
III. Spatial Light Modulators and their Applications for Real-time Optical Processing	20
IV. New Architectures for Optical Processing	30
V. Conclusions	32
References	35
Appendix A - High-Gamma Microchannel Spatial Light Modulator	37
Appendix B - Variable-Gamma Spatial Light Modulator	67
Appendix C - A Comparison of the Hughes and the Loral LCLV Characteristics	195
Appendix D - Selected Publications Resulting from this Contract	267
Appendix E - New Architectures for Optical Processing	334

EXECUTIVE SUMMARY

In the field of real-time optical image processing, an essential optical component is the spatial light modulator (SLM). SLMs can be used for inputting 2-D images in real-time, for forming spatial filters, and for recording computer generated holograms. Before the invention of the SLM, silver halide films were used to perform these functions, but the processing was done in a darkroom in a non real-time mode, i.e., the film had to be developed and fixed before it was useful.

Although film has many advantages, such as its high resolution (~ 2000 lines per millimeter) and proper contrast for optical processing, the time-delay resulting from the darkroom process renders the phrase "processing with speed of light" totally meaningless. Therefore, a major objective of the research reported here is to determine how one can do real-time optical processing through the proper development and application of SLMs.

Optical processing may be classified into two types, according to the relationship between the input and the output. If the output is representable as a linear superposition of the input, the processing is said to be linear.

Linear optical processing has been known for several decades. An example is the removal of additive noise through a linear spatial filtering operation. In contrast to linear processing is nonlinear processing. For example, if the noise and signal are multiplied in the input, they can only be separated via a type of nonlinear processing called logarithmic filtering.

Nonlinear processing was explored in the late 1970's. One of the most successful techniques for nonlinear processing is the halftone method. In this method, the image is first copied through a halftone screen onto a high-gamma (high contrast) film, in a manner similar to that of the printing process. The developed film is called the halftone transparency. When the halftone transparency is placed at the input of a coherent optical processing system, many interesting phenomena may occur at the output plane. These include the level-slicing and equi-density plot of the input, and the pseudo-coloring of a black-and-white picture. Subtle differences in gray levels in the input can be detected via these techniques. The applications of these

techniques for industrial inspection, medical diagnosis, natural resource identification, and military surveillance are apparent.

The final report has documented the essential findings of the research focused on real-time nonlinear optical processing. The report is divided into three parts, all sharing the common objective of developing a real-time processing technology. The first part is the characterization and development of SLMs. The second part is the design, analysis, and fabrication of specific halftone screens suitable for optical processing. The third part is the performance of real-time nonlinear processing and the research and development of new optical processing architectures based on the progress made in the first two parts. More details of the results obtained in these three parts are summarized below.

(1) SLM characterization and development

The micro-channel SLM (MSLM) has been investigated in the following areas: (a) conversion of the standard SLM to a Fabry-Perot MSLM for high-gamma operation, (b) operation of the MSLM in a closed-loop feedback configuration, (c) operation of the MSLM in the electron deposition/grid stabilized mode, and (d) operation in the real-time hard-clipped mode. It has been found that the MSLM can achieve very high contrast (high gamma). However, as the MSLM device available for our experiment had limited spatial band-width product and lack of uniformity, other SLMs, such as the Hughes LCLV, have also been characterized.

The operation of the Hughes LCLV is described below. The birefringence of the liquid crystal (LC) is controlled by a writing beam incident on the writing side and the voltage across the photoconductor and the LC layers. In the reflecting mode of operation, the output is reflected from the reading side of the device through the polarization modulation by the liquid crystal layer. It has been found that the gamma of the SLM may be increased from 1 to 1.7 when the LCLV is biased at lower frequencies and lower voltages than those suggested by its manufacturer for normal operation. However, in this mode, the speed of operation is reduced. Cascading of two LCLVs may provide enough high gamma for nonlinear processing.

For the inputting requirement, inexpensive electronically addressable liquid crystal televisions (LCTVs) have been investigated for optical processing. The LCTV, after proper modification, can be used as an SLM for many optical processing applications, such as pattern recognition, pseudo-color encoding, logic operations, and computer-generated holograms. In its transmission mode, the LCTV SLM can modulate both the amplitude and phase of an input via polarization rotation of the liquid crystal controlled by the applied electric field.

(2) Halftone screen design, analysis, and fabrication

As mentioned, the halftone screen technique can be used for operations such as logarithmic transformation, exponentiation, level slicing, multiple isophote generation, quantization, pseudo-color encoding, A-D conversion, image subtraction, and multiple image storage. All these techniques have been demonstrated in the non-real-time mode. In this part of the research, the pulse-angular-frequency modulation principle has been used for the design of the screen. In the fabrication process, one-dimensional masks are fabricated first, and 2-D screens are then made from the 1-D masks. The expected output at the Fourier plane has to be mathematically analyzed. The 2-D screens can be used for histogram measurement, density slicing, and density-based artificial stereo transformation.

(3) Real-time nonlinear processing and new architectures

The specifically designed halftone screens have been used with an input image in a coherent optical processing system with an MSLM as the nonlinear thresholding device to achieve the nonlinear processing operation. It has been discovered that real-time halftone effects can be obtained through this system. Hard dots of various sizes have been yielded. The only problem is that the MSLM has a limited region of uniform aperture and, therefore, the total SBP is not sufficient for further spatial filtering operations.

Three ideas have been conceived for solving this problem. One is to continue to develop MSLMs which would achieve high-gamma high SBP results. The second is to use the Hughes LCLV in a cascaded mode. The third is to utilize the high contrast capability of a photorefractive crystal, such as BaTiO₃, for the thresholding operation.

In the investigation of real-time processing, new architectures for future investigation have been developed. These include:

- (a) array architectures for edge detection feature extraction and for median/rank order filtering used in image enhancement;
- (b) array architectures using artificial neural networks (ANNs) for image restoration and pattern recognition;
- (c) array architectures for Kalman filtering for least square image restoration/reconstruction.

In addition, a proposal on multi-sensory neural processing, based on the optical implementation of the adaptive linear neural (ADALINE) system and the multiple ADALINE (MADLINE) system, has been written. The objective is to use these neural systems to perform logic functions and, eventually, distortion invariant pattern recognition.

I. INTRODUCTION

In performing real-time optical processing, an essential optical component is the spatial light modulator (SLM). The SLM's are necessary for the provision of real-time input, real-time processing, and spatial filtering with computer-generated holograms or joint transform filters. In performing non-linear optical image processing, an optical component is needed to map an input to an output with a nonlinear function. Among all the devices that can provide such nonlinear mapping, it has been shown that the halftone screen, which has been well known in the printing industry, is a highly versatile and flexible component. A major objective of this research is to evaluate and develop real-time nonlinear optical image processing through the incorporation of spatial light modulators with specifically designed halftone screens.

Another major objective of the research is to use existing SLM's to perform real-time image processing. In addition, by the end of the contract, it is expected that new real-time optical processing architectures will have been discovered.

We have performed work focused on the above-stated major objectives. In the spatial light modulator area, the microchannel spatial light modulator (MSLM) has been developed through a sub-contract with the Massachusetts Institute of Technology (MIT). The goal is to achieve high contrast (i.e., high gamma) operation in the MSLM for optical thresholding in real-time which is important for both non-linear image processing and the recently developed optical neural networks. Other SLM's that we investigated include the newly discovered liquid crystal television SLM's, and the Hughes and Loral (Xerox) liquid crystal light valves. We have characterized these SLM's with respect to their aspects of applications in non-linear image processing.

In the halftone screen part, we have analyzed the design and performance of a special screen which is able to make pulse-frequency modulation and analysis of an input image. Optical methods have been devised for making this type of screen and examples are provided.

In real-time halftone nonlinear processing, we have used the halftone screen in combination with the MSLM and showed that halftone dots can be obtained in real-time. The MSLM, due to its limited number of SBP and resolution, can only provide localized halftone modulation. In order to do larger

image halftoning, we have also experimented with the Hughes LCLV in its high-gamma mode. We recommend for future research that the high gamma nature of photorefractive crystals be explored for the thresholding operation.

Finally, we have suggested as an extension of real-time optical processing, the optical implementations of multi-sensory adaptive neural processing for linear and nonlinear data separation, logic operations, heteroassociative mapping for data reduction, and distortion invariant pattern recognition. The new architecture has many unique features which include the elimination of spurious states and the addition of the controllability of basins of attraction for generalization. More importantly, we propose to use the various SLM's such as the MSLM, LCTVSLM, and Hughes' LCLV for the inputting and thresholding operation in real-time neural processing.

II. HALFTONE TECHNIQUES FOR SPATIAL PULSE-FREQUENCY MODULATION

Optical coding or modulation techniques have been proven useful for many image processing applications. The commonly seen techniques include linear modulations of an image by using Ronchi rulings¹ or sinusoidal gratings,² and spatial pulse-width modulation by contact screens.³ Based on these techniques, operations such as logarithmic transformation,⁴⁻⁶, exponentiation,^{5,6} level slicing,^{5,7} multiple isophote generation,^{8,9} quantization,¹⁰ pseudocolor,¹¹ A-D conversion,^{8,12} subtraction,^{3,13} and multiple image storage¹ can be realized. The one-dimensional (1-D) optical pulse-frequency modulation technique has been developed by Indebetouw in achieving pseudocolor¹⁴ and by Soffer, et al in achieving real-time level slicing through the variable grating mode liquid crystal (VGMLC).¹⁵ In addition, Schneider¹⁶ has made a special contact screen for pulse azimuth modulation.

We have investigated a special two-dimensional (2-D) optical pulse frequency modulation technique that can be used to achieve histogram measurement,¹⁷ density slicing,¹⁸ density-based artificial stereo,¹⁹ and gray level image enhancement.²⁰ We have introduced angular and spatial-frequency transmittance variations into the cell patterns of the 2-D contact screens.

In the following, an analysis of the screen is shown with the illustration of an example. A photograph of the screen made by the technique is also presented.

2-1 DESIGN AND ANALYSIS OF THE SPECIFIC SCREEN

We present a theoretical analysis of the halftone technique with an illustrative example and experimental result as follows.

2-1-1 Screen Design

First, a one-dimensional contact screen is fabricated²¹ with its transmittance pattern¹¹ of its unit-cell $T^s(x)$ shown in Fig. 1. The modulation process is accomplished by contact-printing the original image through the screen onto a (positive or negative) high (infinite to be ideal) gamma film.

Assume that the frequency of the contact screen is sufficiently high so that within the region of any one unit-cell of the screen the transmittance T of the image transparency is constant. Moreover, we define $T^{th} = I^{th}/I_o$ where I^{th} is the threshold intensity of the high gamma film and I_o is the intensity of the uniform source of illumination.²² If the image transmittance (of a transparency) $T < T^{th}/T_1$, the corresponding unit-cell region of the developed film will have a transmittance of nearly 1 in case a negative film is used. This corresponds to a zero spatial frequency, i.e., $f_x = 0$. Likewise, for $T^{th}/T_2 > T \geq T^{th}/T_1$, $f_x = 1/X$; for $T^{th}/T_3 > T \geq T^{th}/T_2$, $f_x = 2/X$; . . . and for $T \geq T^{th}/T_4$, $f_x = 8/X$. Thus the image in the developed film is frequency-modulated.

2-1-2 Optical Image Processing - intensity histogram; image enhancement; pseudo-color; and artificial stereo.

If the developed film is a transparency, it can be placed at the input plane P_1 of a conventional coherent optical image processing system as shown in Fig. 2. At the Fourier plane of the system, the diffraction spot at spatial frequency coordinate $u = f\lambda 1/X$ contains the image information where $T^{th}/T_2 > T \geq T^{th}/T_1$, where f is the focal length of the Fourier transform lens L_1 and λ is the wavelength of point source S . Similarly, the spot at $u = f\lambda 2/X$ contains the image information of $T^{th}/T_3 > T \geq T^{th}/T_2$, etc. Several experiments can then be performed. First, after the light intensities of these spots are measured, a histogram of the original image can be obtained in a straight-forward manner. Second, if different neutral density filters are placed behind those spots and all other diffraction orders are stopped, the output image on the plane P_2 will be changed with respect to the input image.

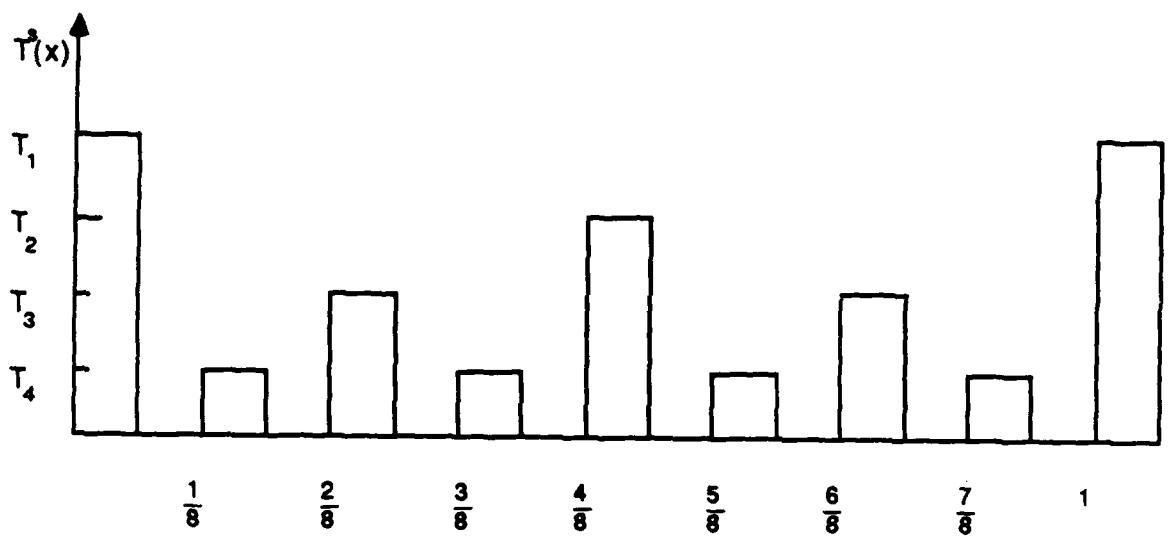


Figure 1. The transmittance of a unit-cell of a specific contact screen for pulse-frequency modulation

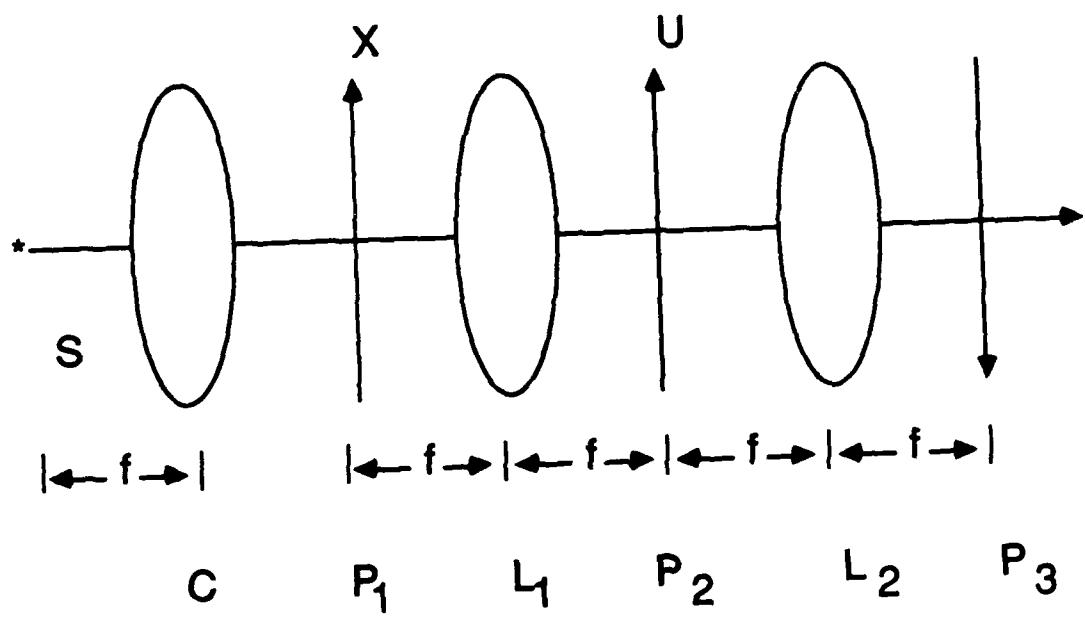


Figure 2.

The coherent optical image processing system

For example, if the density filters have lower densities for the portions with higher T , the image contrast will be enhanced. Third, if a white light source is used instead of the monochromatic source, and a set of color filters are placed at the Fourier plane to selectively pass certain colors, a pseudo-colored image can therefore be obtained. Finally, if a set of optical wedges with different wedge angles are placed behind these diffraction spots, artificial stereo pictures can then be obtained.

2-1-3 Cross-Talk Elimination with Specific Screen Design

It should be pointed out that owing to the cell pattern shown in Fig. 1, the diffraction order with $f_x = 2/X$ overlaps with the second harmonic of the diffraction order with $f_x = 1/X$. The overlapping becomes more serious at higher frequencies. As a consequence, crosstalk among different transmittances in the image would occur.

To avoid the crosstalk problem, a new two-dimensional contact screen with typical cell patterns, e.g., as shown in Fig. 3 should be designed. In Fig. 3, a unit cell consists of three columns denoted by A, B and C, where T_1 , T_2 , T_3 , and T_4 in the parallelograms represent the corresponding local transmittances. Such a screen can be produced by exposing a high-resolution fresh low-gamma film covered with two masks; one of which is a periodic 1-D thin-slit mask (say Mask No. 1) with period X and slit-width $X/3$. The other one is also a periodic 1-D thin-slit mask (say Mask No. 2) with period X' (which may or may not be equal to X) and slit-width $X'/2$. In obtaining the patterns shown in Fig. 3, six exposures should be made. The first exposure is for getting T_1 in the " T_1 " parallelograms in columns marked by A. During the exposure, the slits on the two masks are placed perpendicular to each other. The second exposure is for T_2 in the " T_2 " parallelograms in column A. It is made after mask No. 2 is shifted by a distance $X'/2$ along the direction perpendicular to its slit lines. A third exposure is made after mask No. 1 is shifted a distance $X/3$ perpendicular to its slit lines and mask No. 2 is rotated by an angle θ_B , and T_2 in the " T_2 " parallelograms of column B can be recorded. The processes in getting the transmittances in the other parallelograms are similar to the steps described above.

When the screen with cell patterns shown in Fig. 3 is used for modulating an original image, a frequency-modulated image can then be obtained. In this case, however, the differences among various frequencies are not their

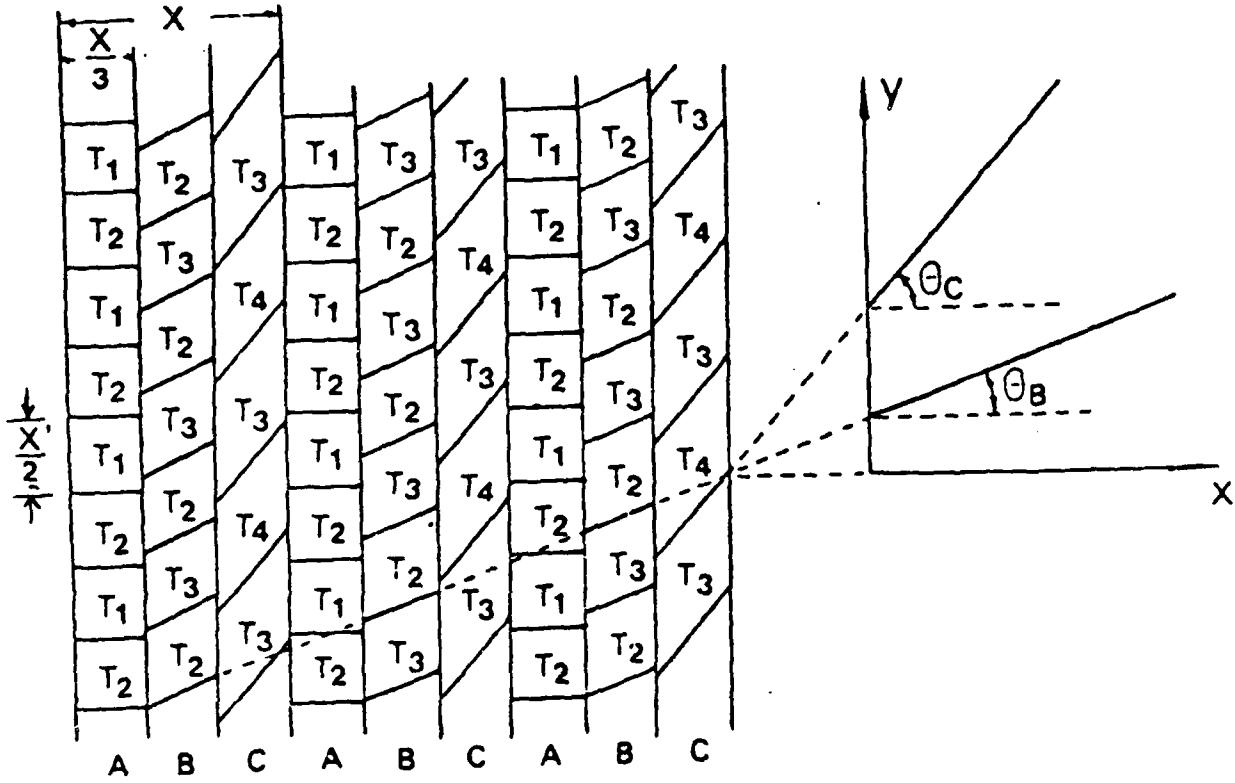


Figure 3.

The transmittance distribution in several cell patterns of a two dimensional pulse-frequency modulation screen

moduli but their azimuths. The cross-talk problem can be avoided if the separations among the frequencies of the masks are sufficient and the angle of the rotation between the two masks are adequate. The zero and first order of the Fourier spectrum of the modulated image would look like that as shown in Fig. 4, where u-axis and v-axis are the spectral coordinate axes corresponding to x, y axis on the modulated image. The spectrum spots denoted by "O", "x", and "□" are corresponding to columns A, B, and C and relate to the ranges of transmittances, $T^{\text{th}}/T_1 \leq T < T^{\text{th}}/T_1$, T^{th}/T_3 , and $T^{\text{th}}/T_3 \leq T < T^{\text{th}}/T_4$ respectively. All the diffraction orders except the ones located on the u-axis can be used for separately modifying the different transmittance components of the original image without suffering from any cross-talk. In general, more transmittance levels can be discriminated or frequency-encoded if there are more columns in a unit cell as shown in Fig. 5.

2-1-4 Mathematical Analysis

The simple example described above is only for the illustration of the basic idea of the pulse-frequency modulation technique. Details of a more general 2-D analysis are presented below.

We assume that the general screen function $f(x,y)$ can be illustrated in Fig. 5, where X is the period of a first slit mask. The slits in the mask are assumed to be oriented along the y-axis. We further assume that the number of distinguishable gray levels is N.

We may consider the whole screen pattern as a sum of N sets of periodical strips. The n^{th} set can be expressed by $p_n \times \gamma_n$, where p_n is a vertical grating as shown in Fig. 6 with its period x and its slit-width X/N , the positions of the slits for p_n relative to those for p_1 are shifted along x axis by a quantity $(n-1) X/N$; γ_n is an inclined grating made by a second slit mask as shown in Fig. 7. Its period is X' . The angle of inclination is θ_n . Within any one grating period, there are two different transmittance levels, T_n and T_{n+1} , each of which occupies one half width of the period. We have

$$f(x,y) = \sum_{n=1}^N p_n(x,y,) \times \gamma_n(x,y) . \quad (1)$$

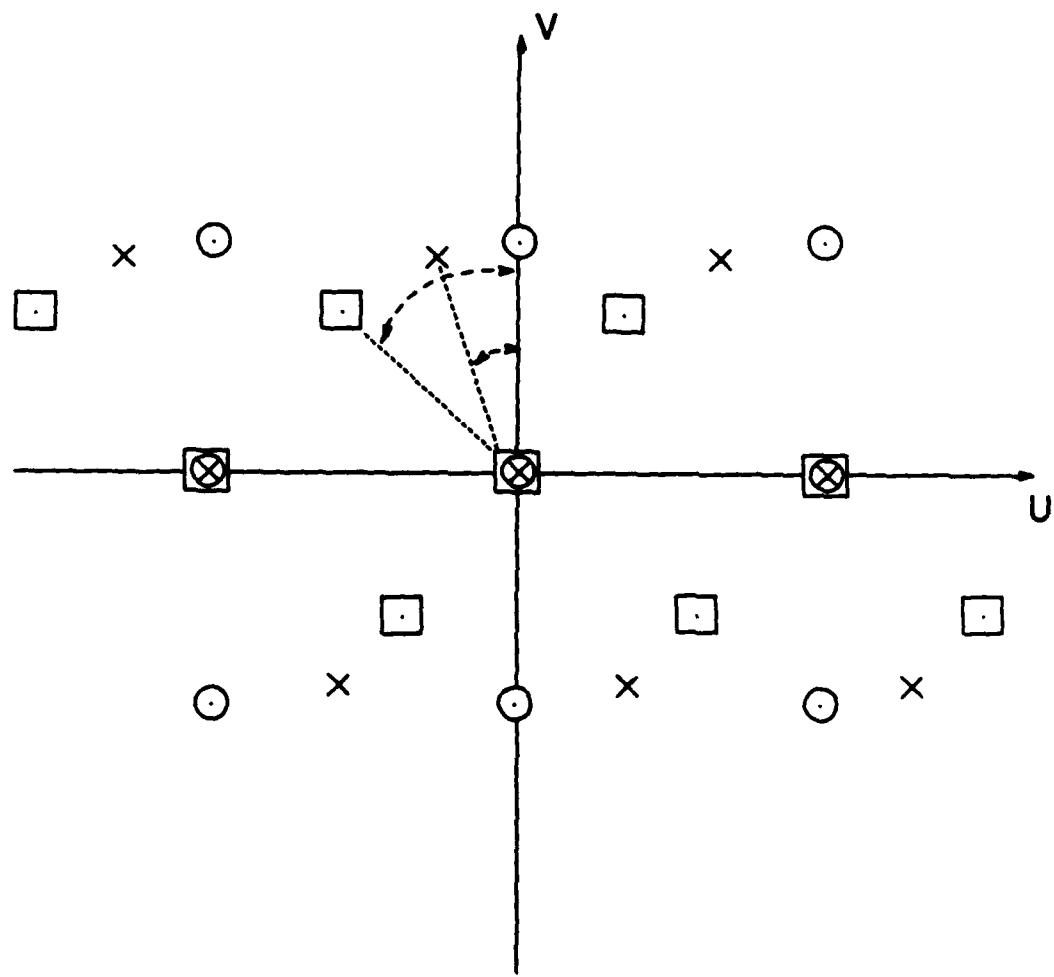


Figure 4. The first and zero-order spectra of the screen pattern of Figure 3

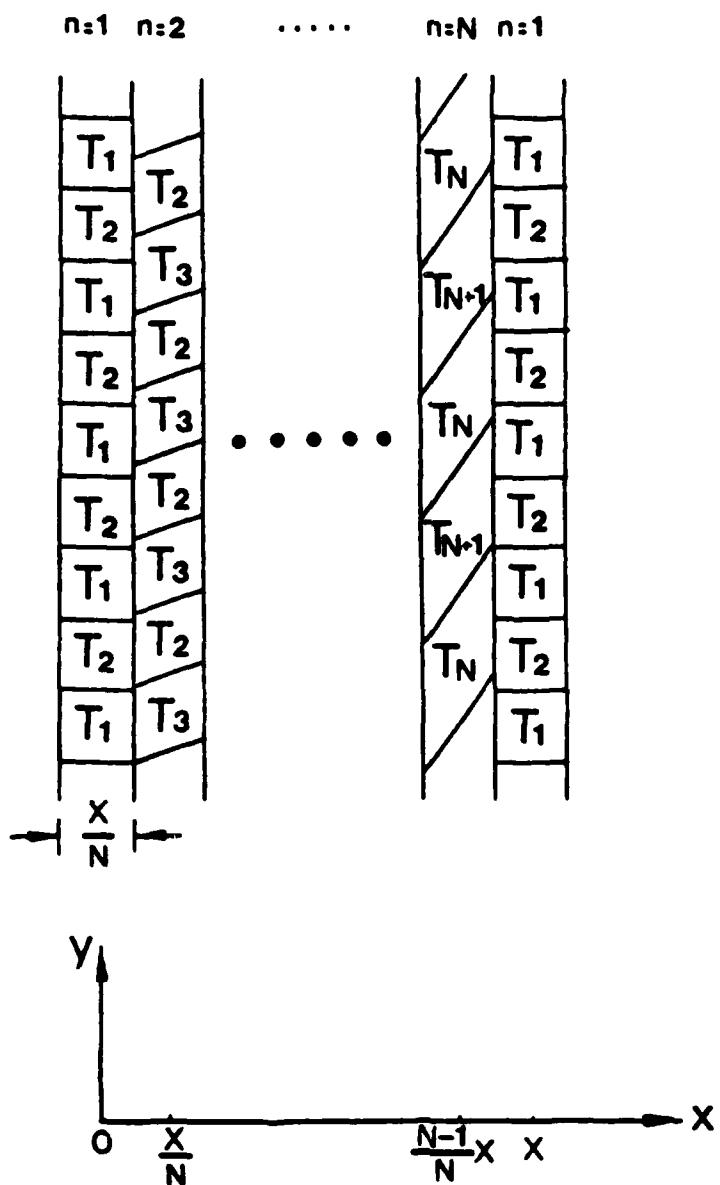


Figure 5. A diagram for illustrating the cell pattern of a typical general 2-D pulse-frequency modulation screen

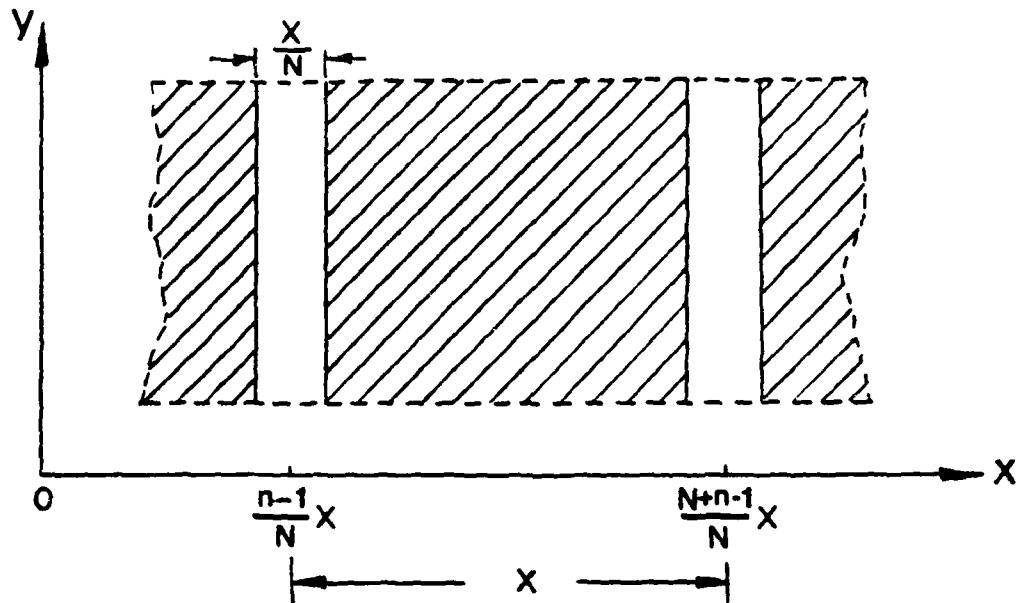


Figure 6 A vertical grating with its period X and its slit-width X/N
made by a first slit mask

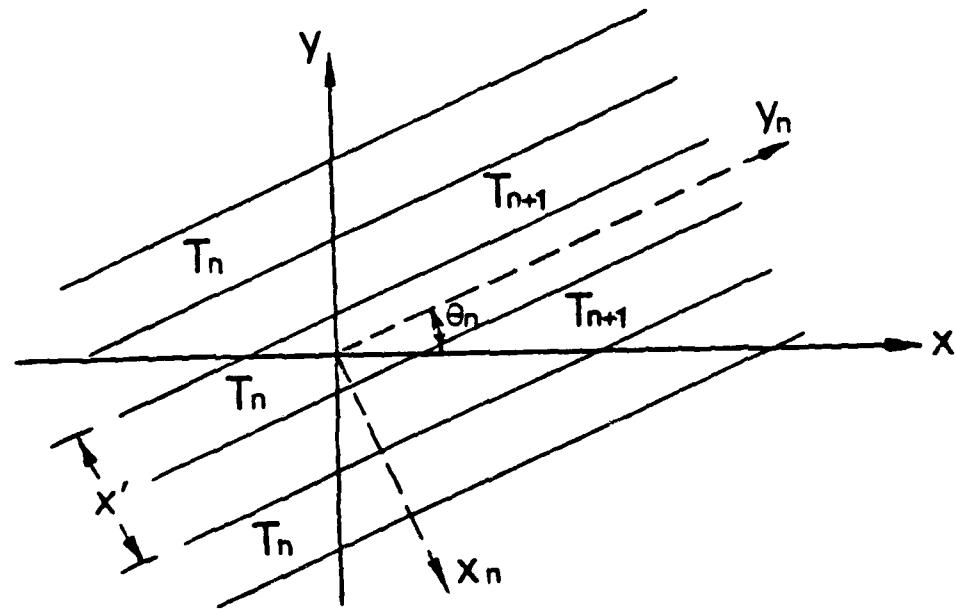


Figure 7. An inclined grating with a period X' angle of inclination θ_n
made by a second slit mask

Notice that

$$p_m(x,y,) \cdot p_n(x,y,) = 0 \text{ for } m \neq n \quad (2)$$

at any point $(x,y,)$. The Fourier transform of $f(x,y)$, $F(u,v)$ is

$$F(u,v) = \sum_{n=1}^N p_n(u,v) * R_n(u,v) \quad (3)$$

where p_n and R_n are the Fourier transforms of p_n and r_n respectively, $*$ denotes convolution, and u,v are spatial frequencies corresponding to x,y .

The mathematical expressions of the functions involved are described as follows:

(a) $p_n(x,y)$ and $P_n(u,v)$

$$p_n = \text{rect}\left(\frac{x - \frac{n-1}{N} X}{\frac{X}{N}}\right) * \sum_{\ell=-\infty}^{+\infty} \frac{1}{X} \delta\left(-\frac{x}{X} - \ell\right) \delta(y) \quad (4)$$

For simplicity, we will hereafter write

$$\sum_{\ell=-\infty}^{+\infty} \quad \text{as} \quad \sum_{\ell}, \quad \sum_{x=-\infty}^{+\infty} \quad \text{as} \quad \int_x, \text{ etc.}$$

The 2-D Fourier transform of P_n is:

$$\begin{aligned} P_n(u,v) &= \frac{X}{N} \text{sinc}\left(\frac{X}{N} u\right) e^{-j2\pi u\left(\frac{n-1}{N}\right)X} x\delta(v) \times \sum_{\ell} \frac{1}{X} \delta\left(u - \frac{\ell}{X}\right) \\ &= \frac{1}{N} \sum_{\ell} \text{sinc}\left(\frac{\ell}{N}\right) e^{-j2\pi \ell\left(\frac{n-1}{N}\right)} \delta\left(u - \frac{\ell}{X}\right) \delta(v) \end{aligned} \quad (5)$$

(b) $r_n(x,y)$ and $R_n(u,v)$

We first express r_n as shown in Fig. 8 in (x_n, y_n) coordinates by

$$r_n'(x_n, y_n) = r_n(x, y) \quad (6)$$

It is obvious that:

$$\begin{aligned} & r_n'(x_n, y_n) \\ &= (T_n \operatorname{rect}(\frac{x_n}{X'/2}) + T_{n+1}[\operatorname{rect}(\frac{x_n}{X'}) - \operatorname{rect}(\frac{x_n}{X'/2})]) * \sum_m \frac{1}{X'} \delta(\frac{x_n}{X'} - m) \delta(y_n) \\ &= ((T_n - T_{n+1}) \operatorname{rect}(\frac{x_n}{X'/2}) + T_{n+1} \operatorname{rect}(\frac{x_n}{X'})) * \sum_m \frac{1}{X'} \delta(\frac{x_n}{X'} - m) \delta(y_n) \end{aligned} \quad (7)$$

The Fourier transform can also be represented in spatial frequency domain (u_n, v_n) by

$$\begin{aligned} R_n(u, v) &= R'_n(u_n, v_n) \\ &= [(T_n - T_{n+1}) \frac{X'}{2} \operatorname{sinc}(\frac{X'}{2} u_n) + T_{n+1} \cdot X' \operatorname{sinc}(X' u_n)] \cdot \delta(v_n) \\ &\quad \times \sum_m \frac{1}{X'} (u_n - \frac{m}{X'}) \\ &= \sum_m [(T_n - T_{n+1}) \cdot \frac{1}{2} \operatorname{sinc}(\frac{m}{2}) + T_{n+1} \operatorname{sinc}(m)] \cdot \delta(u_n - \frac{m}{X'}) \cdot \delta(v_n) \\ &= \sum_m [(T_n - T_{n+1}) \frac{1}{2} \operatorname{sinc}(\frac{m}{2}) + T_{n+1} \operatorname{sinc}(m)] \delta(u \sin\theta_n - v \cos\theta_n - \frac{m}{X'}). \\ &\quad \delta(u \cos\theta_n + v \sin\theta_n (m)). \end{aligned} \quad (8)$$

The last step is resulted from a coordinate conversion. After rearranging the two δ -functions, we have

$$\begin{aligned} R_n(u, v) &= \sum_m [(T_n - T_{n+1}) \frac{1}{2} \operatorname{sinc}(\frac{m}{2}) + T_{n+1} \operatorname{sinc}(m)] \delta(u - \frac{m}{X'} \sin\theta_n) \\ &\quad \times \delta(v + \frac{m}{X'} \cos\theta_n). \end{aligned} \quad (9)$$

Notice that the function $\text{sinc}(\frac{m}{2})$ has non-zero value only for $m = 0$ and $m = \text{odd integers}$, and

$$\text{sinc}(m) = \begin{cases} 1 & m = 0 \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier spatial frequency spectrum of the screen is analyzed below.

For convenience, let

$$A(\ell, n) = \frac{1}{N} \text{sinc}\left(\frac{\ell}{N}\right) e^{-j2\pi\ell \frac{n-1}{N}} \quad (10)$$

$$B_n(m) = [(T_n - T_{n+1}) \frac{1}{2} \text{sinc}(\frac{m}{2}) + T_{n+1} \text{sinc}(m)] \quad (11)$$

$$\text{and } F_n(u, v) = P_n(u, v) * R_n(u, v) \quad (12)$$

$$\text{Thus, } P_n(u, v) = \sum_{\ell} A(\ell, n) \delta(u - \frac{\ell}{X}) \delta(v) \quad (13)$$

$$R_n(u, v) = \sum_m B_n(m) \delta(u - \frac{m}{X}, \sin\theta_n) \delta(v + \frac{m}{X}, \cos\theta_n) \quad (14)$$

$$\text{and } F(u, v) = \sum_{n=1}^N F_n(u, v). \quad (15)$$

As a consequence,

$$\begin{aligned} F_n(u, v) &= P_n(u, v) * R_n(u, v) \\ &= \iint_{u', v'} \left[\sum_m B_n(m) \delta(u' - \frac{m}{X}, \sin\theta_n) \delta(v' + \frac{m}{X}, \cos\theta_n) \right. \\ &\quad \times \left. \left[\sum_{\ell} A(\ell, n) \delta(u - u' + \frac{\ell}{X}) \delta(v - v') \right] du' dv' \right] \\ &= \sum_{\ell} A(\ell, n) B_n(m) \int_v \delta(u' - \frac{m}{X}, \sin\theta_n) \delta(v + \frac{m}{X}, \cos\theta_n) \delta(u - u' + \frac{\ell}{X}) du' \\ &= \sum_{\ell} A(\ell, n) B_n(m) \delta(u + \frac{\ell}{X} - \frac{m}{X}, \sin\theta_n) \delta(v + \frac{m}{X}, \cos\theta_n) \quad (16) \end{aligned}$$

The positions of the zero and first orders of these spectral components are shown in Fig. 9. The positions and the F_n values of several orders are listed in Table I.

TABLE I. Amplitudes and Positions of the Fourier Spectra of the Screen.

(l, m)	Position (u, v)	F_n
0, 0	0, 0	$A(0, n) B_n(0) = \frac{1}{2N} (T_n + T_{n+1})$
$\pm 1, 0$	$\pm \frac{1}{X}, 0$	$A(\pm 1, n) B_n(0) = \frac{1}{2N} (T_n + T_{n+1}) \operatorname{sinc}(\frac{1}{N}) e^{-j\frac{2\pi(n-1)}{N}}$
$0, \pm 1$	$\pm \frac{\sin\theta_n}{X}, \mp \frac{\cos\theta_n}{X}$	$A(0, n) B_n(1) = \frac{1}{\pi N} (T_n + T_{n+1})$
$\pm 1, \pm 1$	$\pm \frac{\sin\theta_n}{X}, \mp \frac{1}{X}, \mp \frac{\cos\theta_n}{X}$	$A(\pm 1, n) B_n(\pm 1) = \frac{1}{\pi N} (T_n - T_{n+1}) \operatorname{sinc}(\frac{1}{N}) e^{-j\frac{2\pi(n-1)}{N}}$
$0, \pm 2$	$\pm \frac{2\sin\theta_n}{X}, \mp \frac{2\cos\theta_n}{X}$	$A(0, n) B_n(\pm 2) = 0$
$\pm 1, \pm 2$	$\pm \frac{2\sin\theta_n}{X}, \mp \frac{1}{X}, \mp \frac{2\cos\theta_n}{X}$	$A(\pm 1, n) B_n(\pm 2) = 0$

It is easy to find out that just as one expected, all positions of the spectra are located at a set of parallel lines that are inclined to v-axis by an angle θ_n and separated with each other by a distance $1/X$ along u-axis. The locations of the spectra on each of the parallel lines are exactly the same as those of the spectrum of $r_n(x, y)$.

Based on Eq. (15), the whole spectrum of the screen is a summation of all F_n 's ($n=1, 2, \dots, N$). That means all spectral components are located on a group of parallel line sets, and under a careful arrangement, no components of low orders (such as lower than the 7th order, for practical applications, higher orders are useless because of their trivial intensities and first or-

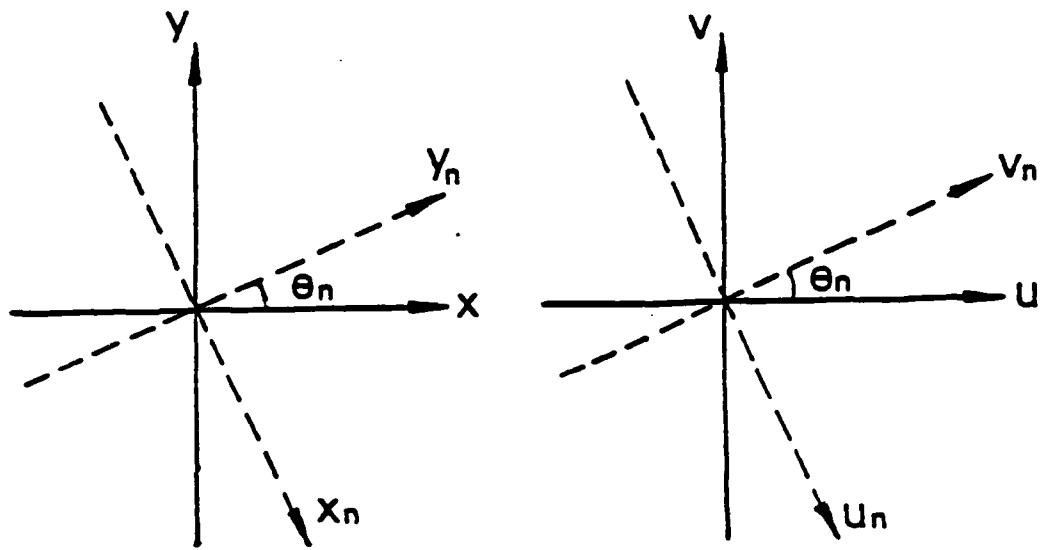


Figure 8.

The coordinates for functions $r_n(x, y)$, $r'_n(x_n, y_n)$, $R_n(u, v)$ and $R'_n(u_n, v_n)$

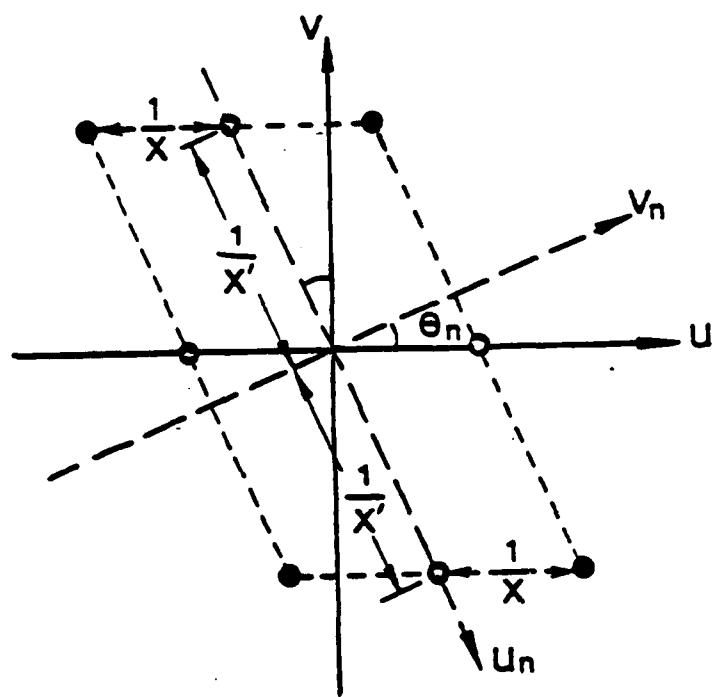


Figure 9.

The positions of the zero and first orders of the Fourier spectra of the screen

ders are enough for most cases) except the zero-order are spatially coincident with others. Figures 3 and 4 show an example with $N = 3$, where the columns A, B and C correspond to $n = 1, 2$ and 3 , respectively, and $\theta_1 = 0$, $\theta_2 = \theta_B$, $\theta_3 = \theta_C$. The locations of the zero and first orders of $F(u, v)$ as shown in Fig. 4 can be determined based on Eqs. (12), (13) and (14).

The spectrum of the pulse-frequency-modulated image is analyzed as follows. After an input image is contact-printed onto a high gamma film through the above-described screen, a pulse-frequency modulated (or encoded) image (transparency) is obtained. The spectrum of the encoded image at the Fourier plane of a Fourier transform lens is similar to but different from the spectrum of the screen in two aspects. (1) Each order is no longer a sharp point-like spot but is somewhat smeared out. The sizes of the smeared spectral islands depend on the size of the smallest area on the original image where the transmittance is between T_n and T_{n+1} if a monochromatic point source is used in the optical transform system. A broadening of the size and band width of the light source will further spread the size of the island. (b) The energy distribution among the various orders of the spectrum of the encoded image is different from that of the screen. This is due to the following reasons: (i) The encoded image is binary but the screens have many gray levels. Thus the spectral coefficients of R_n are different for the two cases. (ii) The relative energy ratios among different n value is also different for the two cases. For the screen, the distribution is fixed, but for the encoded image, it depends on the gray level distribution (histogram) of the original image. For example, if nowhere of the original image has a transmittance value between T_n and T_{n+1} , then the corresponding spectrum order of the encoded image will disappear.

Although the useful spectral orders of the encoded image are smeared, their centers are definitely separated in space. So, the total overlap for the spectral components of different gray levels happening in the 1-D screen case, as we mentioned above, does not happen in the 2-D screen case. Consequently, a 2-D screen can be designed to separate the spectral spots of its encoded image and eliminate cross-talk. To achieve the optimum screen design, factors such as the size and bandwidth of the light source, the number of distinguishable gray levels, and the spatial frequency of the original image should be carefully considered.

Experimental Demonstration

To fabricate the specific halftone screen, a Ronchi Ruling is used to fabricate the mask to be placed on a Kodak 649F glass plate. On the mask is a periodic pattern in both x and y directions. Within each period is a transparent square with opaque background. Each side of the square is $X/6$ inches, where X is the period of Ronchi Ruling. A x-y translation stage was used to move the glass place underneath the mask.

To fabricate a 300 lpi screen, a 300 lpi Ronchi Ruling was used to fabricate the mask on a 4" by 5" Kodak 649 glass plate. On the mask is a periodic line pattern with a period of 1/300 inch in the x direction. The width of the transparent line is about 1/4000 of an inch. A translation stage is used to push the mask which sits on top of the glass plate. The plate was fixed on a platform adjacent to the translation stage.

A translation stage was used to push the mask which sits on top of the glass plate with no gap in between. The place is fixed on a platform adjacent to the translation stage. The Ronchi Ruling remains on top of the mask with no gap in between. A second translation stage is used to push the Ruling to obtain the desired angles.

Six exposures were made, two each with the Ruling positioned at 0, 30, and 60 degrees, respectively. For each angle, after the first exposure, the Ronchi Ruling was pushed by 1/200 of an inch and exposed before it was rotated by 30 degrees.

A photomicrograph of the screen cells that have four different angular orientations and 10 levels of gray in each direction is shown in Fig. 10. Screens of periodicity ranging from 50 lpi to 300 lpi can be made without difficulty. These screens can be made to accomodate SLM's of a variety of resolution limits.

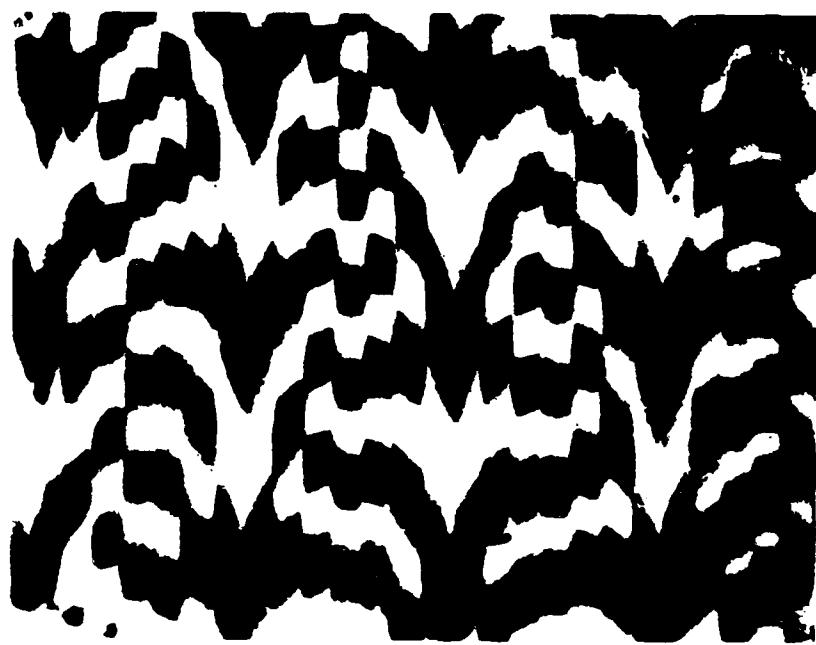


Figure 10. A photomicrograph of halftone screen cells designed for pulse frequency modulation applications

III. SPATIAL LIGHT MODULATORS AND THEIR APPLICATIONS FOR REAL-TIME OPTICAL PROCESSING

We have investigated several SLMs for real-time optical processing.²³ These are the Microchannel Spatial Light Modulator (MSLM), the Hughes and the Loral Liquid Crystal Light Valve (LCLV), and the recently discovered liquid crystal television SLM. For real-time processing, the input film should be replaced by an SLM. The optical thresholding should also be performed by an SLM. These SLMs are described below.

3-1 The MSLM

The standard Microchannel Spatial Light Modulator (MSLM) is a versatile, real-time optical signal and image processing device which exhibits high optical sensitivity and high framing speed. It consists of a photocathode and a microchannel plate (MCP) nearly in focus with an optical planar acceleration grid and an electro-optic crystal plate, as shown in Fig. 11. The electro-optic plate carries a high resistivity dielectric mirror on one side and a transparent conducting electrode on the other.

In the electron deposition write mode, incoherent or coherent write light incident on the photocathode creates an electron image which is amplified by the MCP and focused onto the surface of the dielectric mirror. The resulting spatially varying electric field modulates the refractive index of the electro-optic plate. Thus, the readout light, which makes a double pass through the crystal, is spatially phased or amplitude modulated depending on the crystal cut and readout scheme (polarization or interferometric) employed.

The image is erased by appropriately adjusting the device voltages and flooding the photocathode with light so that electrons are removed from the mirror surface by secondary electron emission. Alternatively, the device can be operated in the reverse mode where a positive-charge image is written by secondary electron emission and erased by adding electrons to the crystal. For Pockels effect crystals, this leads to contrast reversed images.

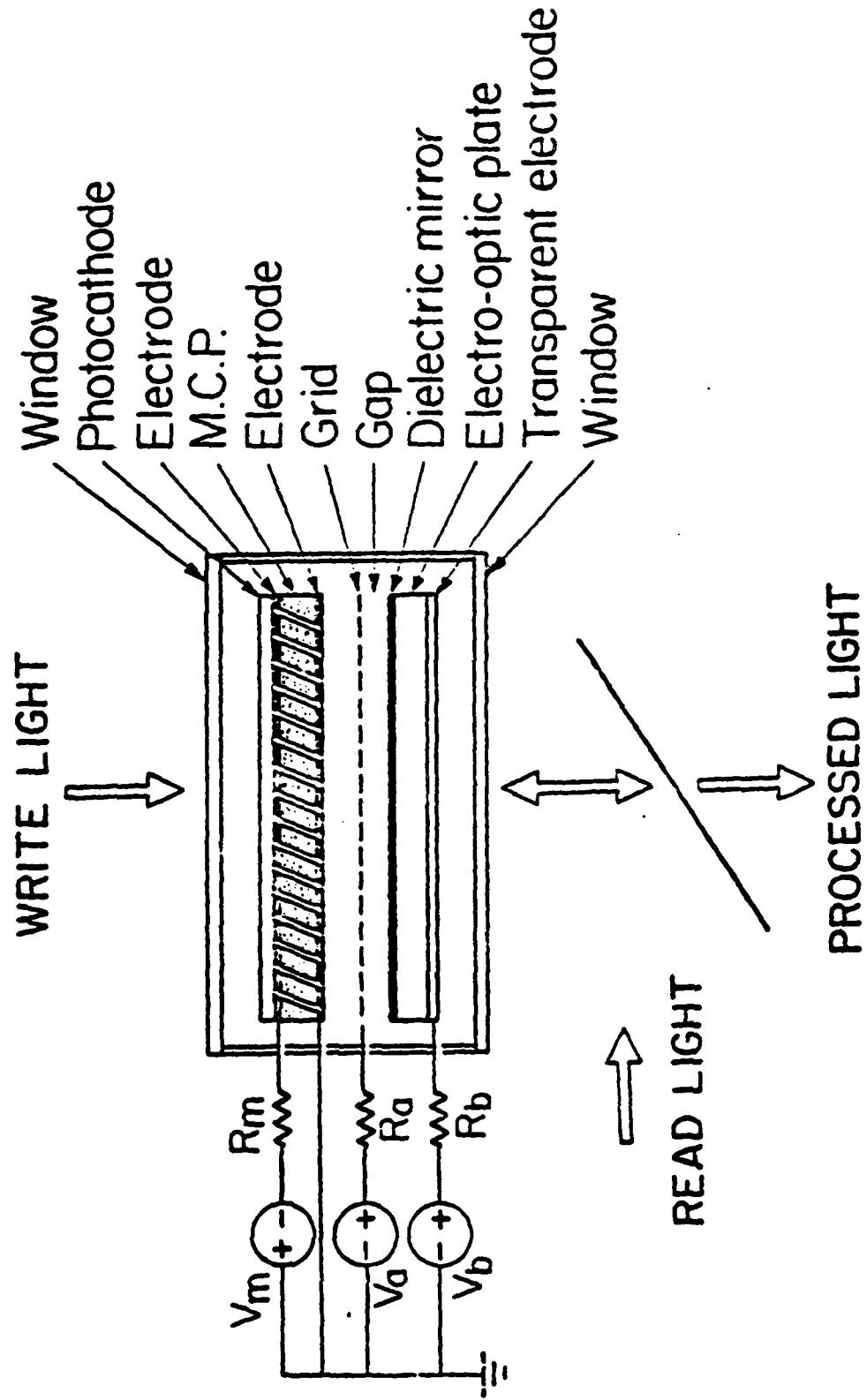


Figure 11.

The Optically Addressed Microchannel Spatial Light Modulator
(Photo-MSLM)

The gamma of the standard device is too low for application in the halftone pulse frequency modulation. To increase the gamma of the MSLM, we need to modify the device by depositing mirrors on both surfaces of the crystal so as to convert it to a spatially tunable Fabry-Perot etalon. A standard dielectric mirror is deposited on the surface that received the electrons, while a conducting mirror is employed on the readout surface of the crystal. The conducting mirror also serves the role of the conducting electrode.

The reason a Fabry-Perot crystal leads naturally to a high gamma device is obvious from an examination of the readout intensity characteristic of the device. This characteristic is shown in Fig. 12. There it is seen that the output intensity becomes a strong function of the input exposure (shown as surface charge density) in certain regions of operation of the device.

With this output characteristic, the device is operated with a bias surface charge density σ_A which biases it at the point A in the operating region when the exposure level is zero. Note also that the device functions like a negative transparency, and that an upper limit must be set on the exposure level to avoid roll over. That is, the device must be operated between the points A and E in Fig. 12.

To determine the gamma of the device, a plot of $\log 1/T$ vs. $\log E$ (where T is the output intensity transmittance for a given exposure E) can be derived from Fig. 12. The slope of this curve will be gamma, γ . That is,

$$\gamma = \log(T_1/T_2)/\log(E_2/E_1) \quad (17)$$

To first order, the incremental surface charge density $\sigma(E)$ deposited on the crystal is proportional to the exposure E.

$$\sigma(E) = (\frac{\eta e G}{h\nu})E \quad (18)$$

where η is the quantum efficiency of the photocathode, e is the electronic charge, G is the gain of the MCP, h is Planck's constant and ν the frequency of the write light. Additionally, the induced phase change $\Delta\phi$ in the crystal is proportional to $\sigma(E)$ for Pockel's effect crystals. The exact expression depends on the type and cut of the crystal employed.

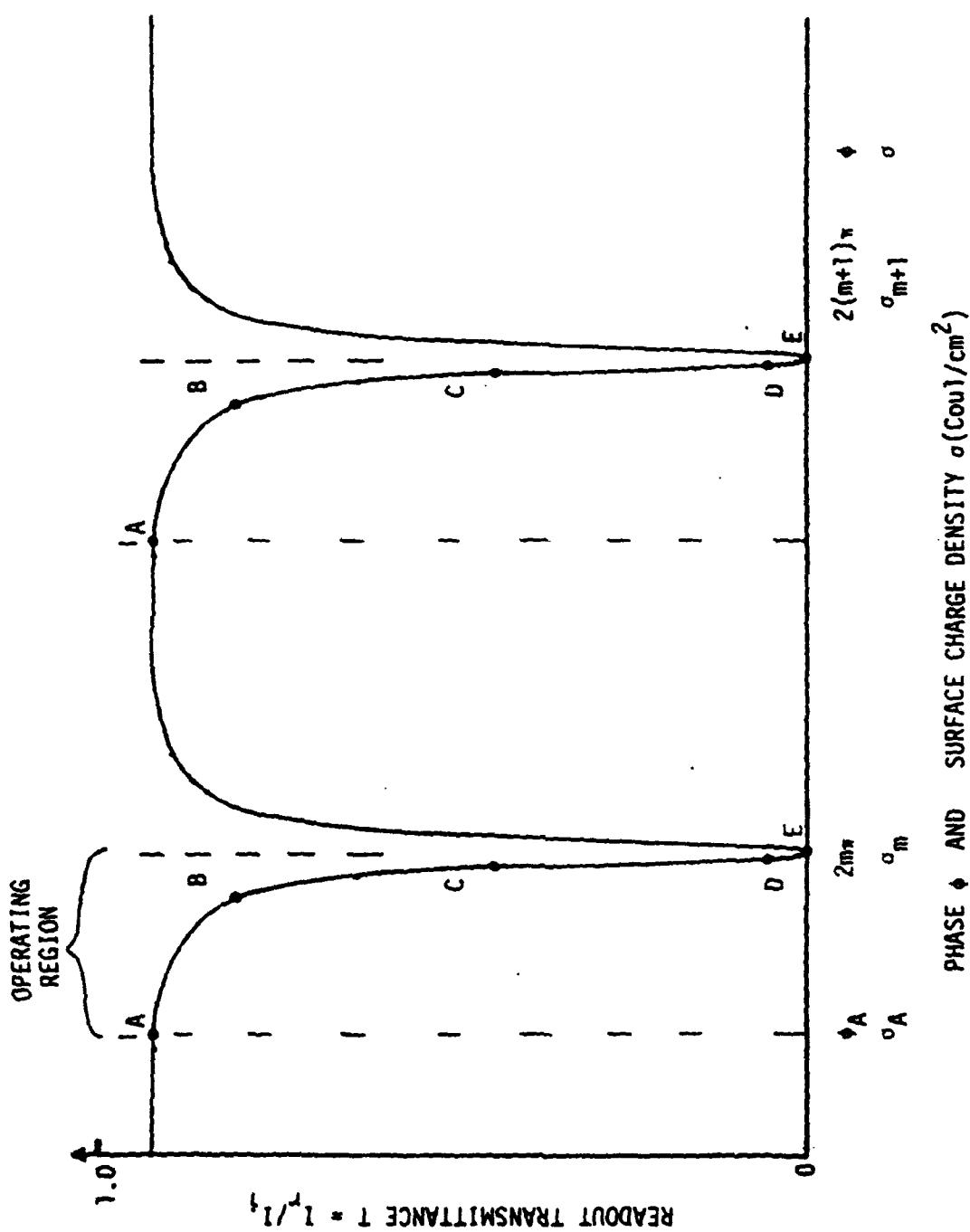


Figure 12. Readout intensity characteristic of a high-gamma Fabry-Perot MSLM

For a Fabry-Perot etalon with surfaces of reflectivity R, the ratio of the total reflected intensity to the incident intensity (transmittance T in MSLM terms) is given by

$$I_r/I_i = T = \frac{4R\sin^2(\frac{\phi}{2})}{(1-R)^2 + 4R\sin^2(\frac{\phi}{2})} \quad (19)$$

Thus the transmittance of the MSLM approaches zero when ϕ takes on multiple values of 2π . Equations (17) through (19) can be solved to yield an approximate theoretical value for gamma within the operation region.

Further details of the MSLM work is described in a technical report entitled, "Bistable Configurations of the MSLM" and a Master's thesis by S. Lau submitted to JPL by Professor C. Warde of the Massachusetts Institute of Technology (MIT). The report and the thesis are attached as Appendix A and Appendix B, respectively.

3-2 The Hughes and Loral LCLV

A commercially available liquid crystal SLM is the Hughes LCLV. The Hughes LCLV has been found useful in many optical processing applications since its introduction in 1970. Another LCLV SLM that was introduced, also in the 1970's, is the Loral EOS device. The latter was still under development. We have measured both the Hughes and the Loral LCLV's for the purpose of comparison. We have found that the two devices are sensitive to different regions of the emission spectra. The Hughes LCLV is more toward 514.5 nm and the Loral is centered in the red region. The reflectivity of the Hughes device is a little higher and so is the writing sensitivity. The operating frequency and operating bias of the Loral LCLV is lower than the Hughes LCLV. However, Loral's LCLV does not have a light blocking layer which may be a problem for some applications. The details of the experimental results are included in Appendix C.

Recently, we have found that by lowering the operating frequency of the Hughes LCLV to a value less than 1 kHz from the manufacturer's suggested 10 kHz, and lowering the bias voltage to around 7 volts, a higher contrast (higher gamma) of the device's operation can be obtained. The measured output in terms of rotation angle versus input writing intensity is shown in Fig. 13.

The gamma estimated under these conditions is around 1.7 in comparison to the 0.8 obtained under normal operating conditions. A corresponding characteristic curve of the Hughes LCLV under normal operating conditions is shown in Fig. 14 for comparison purposes.

Under these conditions, we observed edge-enhancement of continuous tone input, and the results were presented at the OSA SLM Topical Meeting in a paper entitled, "Real-time Optical Edge Enhancement Using a Hughes Liquid Crystal Light Valve"²³. This paper and a collection of publications related to this contract is included in Appendix D.

3-3 The Liquid Crystal Television (LCTV) SLM

One exciting recent discovery in the SLM area is that the LCTV SLM can be modified to serve as an electronically as well as optically addressed SLM. The LCTV is inexpensive and allows a broad range of experimentation with the possibility of scaling up to layered architectures. Its speed is around 30 Hz which is compatible with the existing computer addressing speed. The low contrast and low SBP problems are being investigated by the TV manufacturers and breakthroughs may come forth in the future.

Realizing the importance of the LCTVs for SLM applications to real-time optical processing, we have measured the characteristics of three LCTV SLMs. The pre-prints and reprints of papers on this work are included in Appendix D. These papers contain detailed descriptions of the LCTV SLM and its application.

3-4 Real-time Optical Processing Applications of the MSLM and LCLV

The application of the high-gamma MSLM in combination with a halftone screen for non-linear optical processing was described in a paper entitled, "High-Gamma Spatial Light Modulator for Nonlinear Optical Processing"²⁴ authored by C. Warde, H. Lamela-Rivera, and H. K. Liu, and was published in SPIE

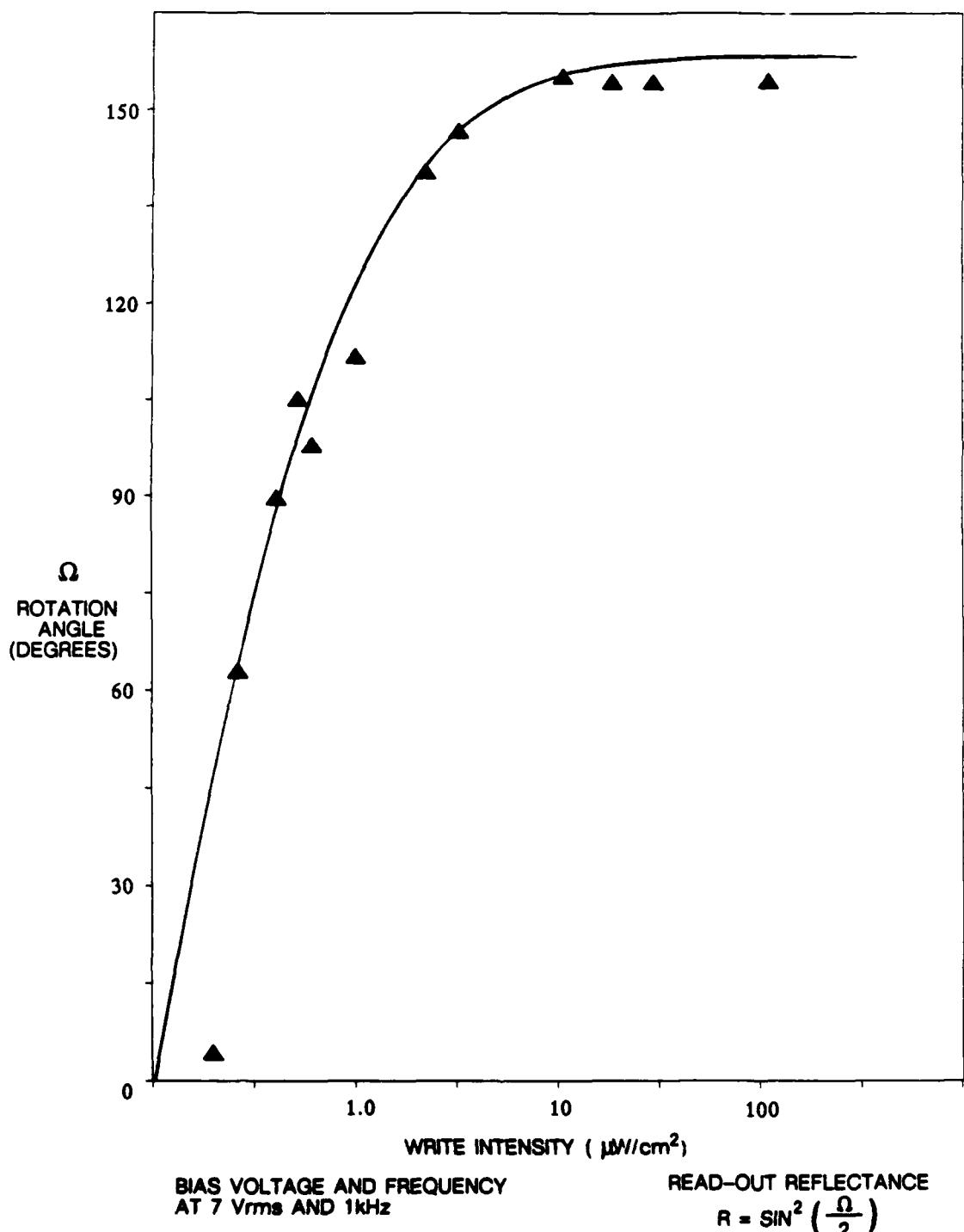


Figure 13.

Polarization rotation angle versus write intensity of the Hughes LCLV under specific bias conditions for high contrast operation

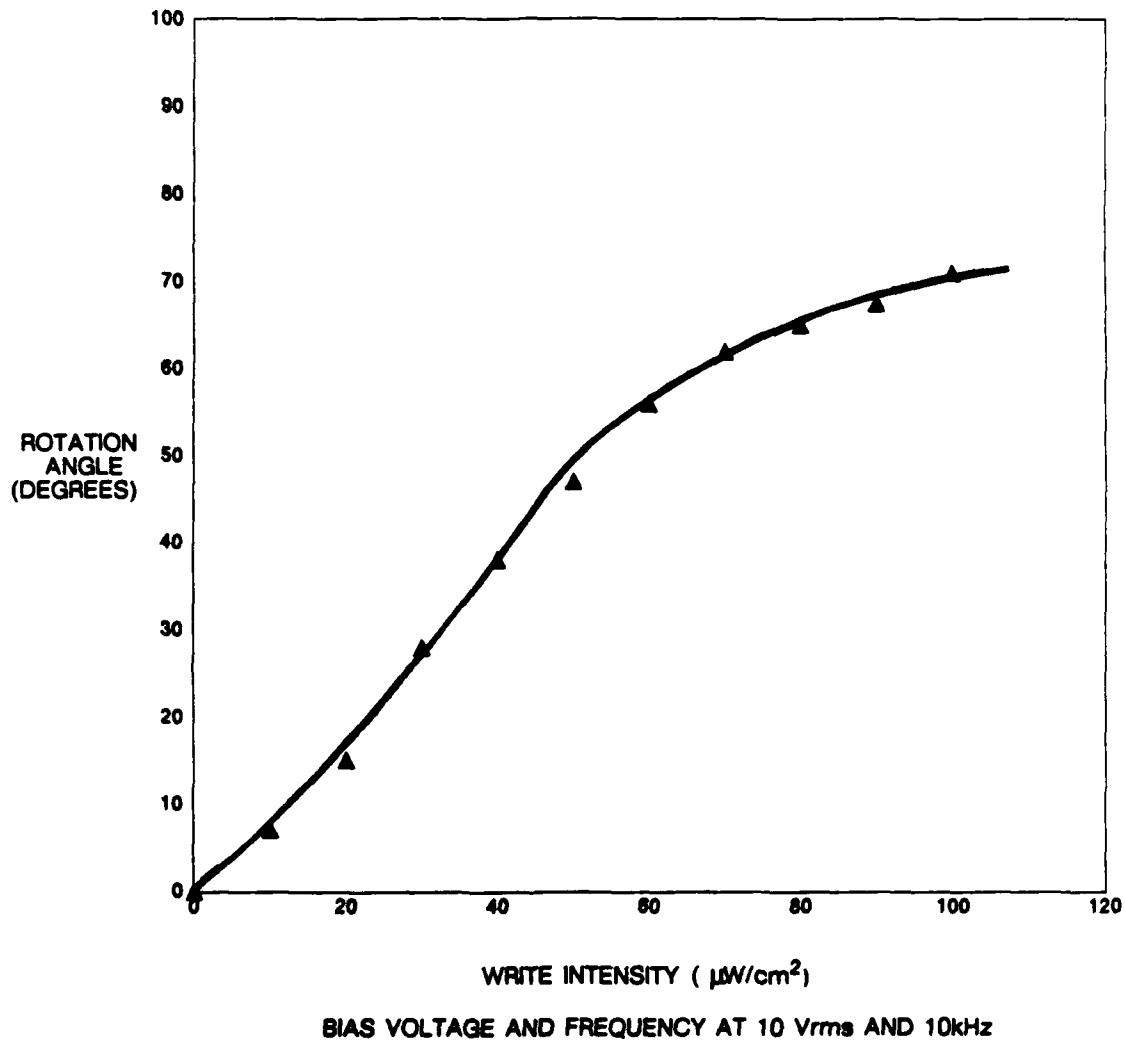


Figure 14.

Polarization rotation angle versus write intensity of the Hughes LCLV under normal operation conditions

Proceeding 613. This paper is included in Appendix D. Further results are included in pages 155-166 of Appendix B. It has been concluded that the halftone effects can be clearly seen from the thresholding of the MSLM. However, SBP of the existing MSLM is not sufficient to be used for image-halftoning. The uniformity of the MSLM needs to be improved.

For this reason, we have also tried to use the Hughes LCLV for the high gamma operation. The set-up of using the LCLV is shown in Fig. 15. Our experimental results indicate that due to the fact that $\gamma = 1.7$ is still not quite sufficient, the halftone effects are not apparent. Two methods can be utilized to improve the gamma. One is through a positive feedback as suggested by A. Sawchuk, but the speed of operation would be slow. Another method is to cascade two LCLVs in series. This set-up is shown in Fig. 16. Since we have only one LCLV at this moment, this method will be tested in the future when two LCLVs are available.

3-5 New Real-time Optical Processing Approaches

A new idea that may also be tried is to use the high gamma thresholding capability of a photorefractive crystal. This is to utilize the spatial light modulation capability of the photorefractive crystal. We plan to try this method in a future research project.

The utility of the LCTV²⁵⁻²⁶ for real-time optical processing is reported in several papers that are also included in Appendix D. We have used the LCTV for writing computer-generated holograms²⁷, pseudo-color encoding for image enhancement²⁸, and associative retrieval neural processing²⁹⁻³¹.

A technique for real-time dynamic range comparison using photorefractive crystals³² and a review on optical subtraction techniques³³ are also presented in Appendix D. These works have contributed to the advancement of real-time optical image processing.

IV. NEW ARCHITECTURES FOR OPTICAL PROCESSING

We have explored the applications of SLMs and real-time optical processing techniques for implementing new image processing architectures. These are divided into two parts. The first part is the study of optical array ar-

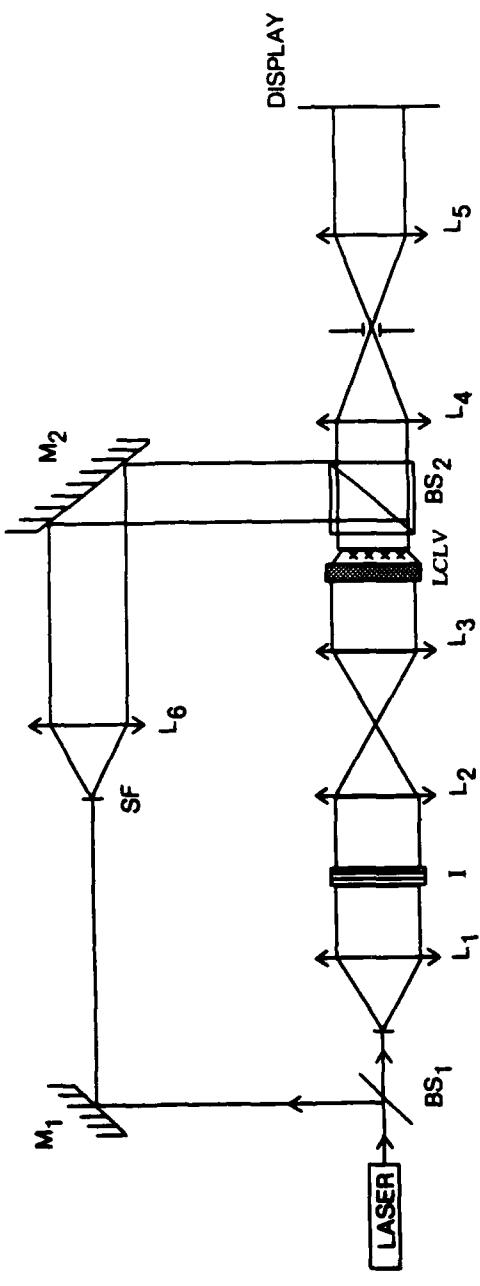


Figure 15. Real-time halftone non-linear processing utilizing the high-gamma operation of a single Hughes LCLV. The input I consists of the input transparency and a halftone screen.

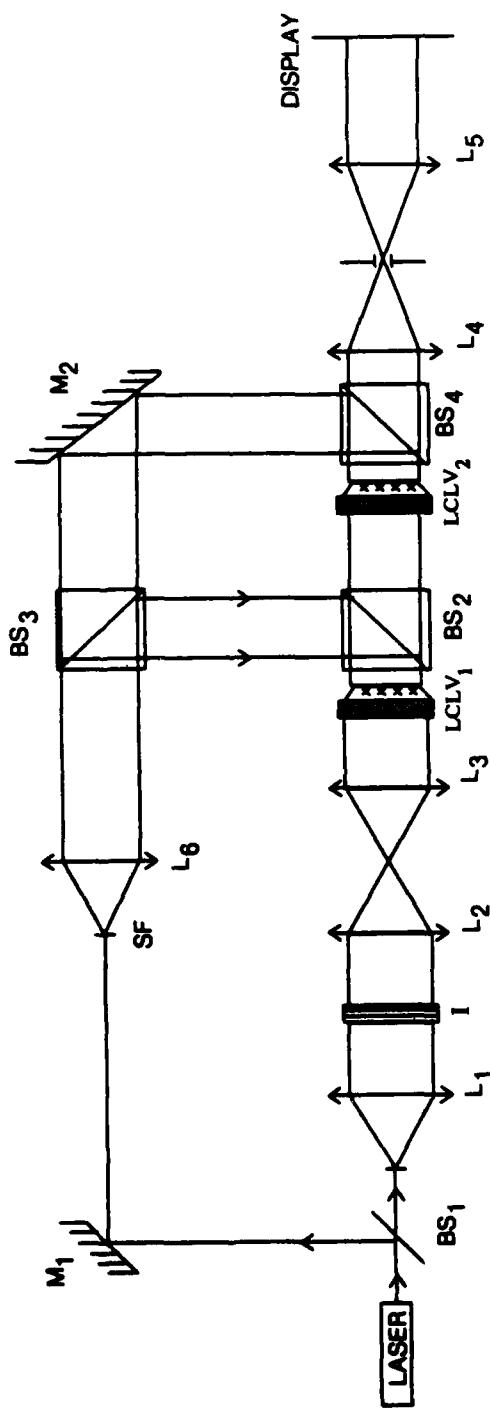


Figure 16. Real-time halftone non-linear processing utilizing the high-gamma operation of two Hughes LCLVs cascaded. The input I consists of the input transparency and a halftone screen.

chitectures for image processing. The second part is the adaptive multisensory neural processing. We describe these two parts briefly below, and refer the interested reader to more details as included in Appendix E.

4-1 Optical Array Architectures

In optical/VLSI development, the current emphasis is on the reduction of overall interconnection complexity and the preservation of parallel and pipelined architecture. There are three types of architectures that may incorporate optical processing. These are (1) array architectures for edge detection feature extraction, and for the median/rank-order filtering used in image enhancement, (2) array architectures using artificial neural networks for image restoration and pattern recognition, and (3) array architectures for Kalman filtering for least squares image restoration/reconstruction. A detailed report by S. Y. Kung on this topic is included in Appendix E.

4-2 Adaptive Multi-sensory Optical Neural Processing

The principle of neural networks is inherently applicable in the processing of multi-sensory information originating from different types of sensors or sources. These data usually need to be digested rapidly to form a coded data knowledge base that can be used for subsequent fast analysis of new "situations" being monitored by the same set of sensors or information gathering sources. An example of multi-sensory, multi-dimensional processing is signal processing based on weather conditions analysis. A multi-channel phase (or vectorial) space concept of the weather condition can be represented by a single phasor in the phase space that contains components of humidity, temperature, and pressure, etc. The phaser moves within the phase space as the weather conditions varies. The multi-channel data can be partitioned into different subgroups of an N-dimensional neuron vector, where N is the total number of neurons, each of which can be "1" or "0" in unipolar binary number representation. Multi-channel spectrogram and sonogram are also problems that can be dealt with by neural networks.

Neural net models offer a powerful methodology to information processing that is iterative, distributive, and nonlinear. One feature of biological neural nets (BNNS) is the maintenance and enhancement of the survivability of a living organism in a fuzzy world where sensory impressions are sketchy. Artificial neural nets (ANNS) enhance survivability through fault-tolerance,

robustness, speed, and adaptability. Survivability implies adaptability through self organization, learning, and fast and accurate decision making. All of these concepts are found extremely important in sociology, economics, and military science for national defense. Neuromorphic computing will make a significant impact on future technology of collecting, aggregating, processing, and evaluating information for industrial and DOD applications.

In the implementation of a multi-channel neural processing network, we must address two basic issues, namely massive interconnectivity and artificial plasticity for self-programming. The massive parallelism and free-space photonic interconnection capabilities of optics should be utilized to resolve these basic problems. For this reason and as a natural extension of the current research topic, we have devoted some effort to the development of a new idea in optical neural processing. We have proposed the investigation of multi-sensory neural processing the investigations of the optical implementations of the basic structure of adaptive linear neurons (ADALINE) for linear and nonlinear separability as originally suggested by B. Widrow. This basic structure would provide learning capability in the network. The basic optical ADALINE unit can be organized to perform AND, OR, and MAJ logic functions. The work will then be expanded to do training in multi-channel and multi-layer adaptive neural processing units that are essential for translation, rotation, scale, and perspective invariant pattern recognition. Details of this processing technique will be submitted to ARO as a research proposal. The various SLMs investigated in this contract can be used for input, memory, and thresholding functions as proposed in the new architecture. The halftone techniques developed are also useful for preprocesing input data for image enhancement and data reduction for neural processing.

V. CONCLUSION

The research on real-time optical image processing was originally inspired by the idea of using a specifically designed spatial pulse-width and pulse frequency modulation halftone screen and a high-gamma micro-channel spatial light modulator for the analysis of an input image in terms of its spatial frequency content. This real-time operation can have many image under-

standings and image-enhancement applications. The design and analysis of the pulse-frequency modulation was completed successfully. In the development of the high-gamma MSLM, it was discovered that the resolution, the SBP, and the uniformity of the SLM still need to be improved before it can be used for non-linear real-time image processing of images of large two-dimensional format. Preliminary experiments, nevertheless have shown that real-time halftone effects are highly visible by using this technique. This indicates the promise of this SLM as the SLM with high gamma. Such a high-gamma SLM will also be needed in neural processing where SGN function thresholding is used in the iteration process.

The LCTV SLM and LCLV SLM investigated in this contract are helpful for real-time optical processing as evidenced by the papers published by this group and others in this country. When the resolution and contrast are improved further by their manufacturers, the LCTV SLM's optical processing applications will be even more broadened.

One important result of this research is that it has opened up possibilities for developing new optical array processors and multi-sensory adaptive neural processing techniques. Such new work ideas capitalize on the unique strength of optics and offer an opportunity for optics to make significant contributions in data processing for future generations of special-purpose computers.

REFERENCES

1. P. F. Muller, "Linear Multiple Image Storage," *Appl. Opt.* 8, 267 (1969).
2. F. T. S. Yu and M. S. Dymek, "Optical Information Parallel Processing: A Technique," *Appl. Opt.* 20, 1450 (1981).
3. D. Z. Zhao, C. K. Chiang, and H. K. Liu, "Quantitative Optical Image Subtraction by Means of Pulse-width Modulation Through One-dimensional Contact Screens," *Optics Lett.* 6, 490 (1981). Also, "Contact Screen Image Subtraction Technique," *Appl. Opt.*, 20, 4234 (1981).
4. H. Kato and J. W. Goodman, "Nonlinear Filtering in Coherent Optical Systems Through Halftone Screen Processes," *Appl. Opt.* 14, 1813 (1975).
5. S. R. Dashiell and A. A. Sawchuk, "Nonlinear Optical Processing: Analysis and Synthesis," *Appl. Opt.* 16, 1009 (1977).
6. M. A. Karim and H. K. Liu, "All Optical Homomorphic Image-Processing System," *J.*, 371 (1982).
7. S. R. Dashiell and A. A. Sawchuk, "Optical Synthesis of Nonlinear Non-monotonic Functions," *Opt. Comm.* 15, 60 (1975).
8. T. C. Strand, "Non-monotonic Nonlinear Image Processing Using Halftone Techniques," *Opt. Comm.* 15, 60 (1975).
9. H. K. Liu, J. W. Goodman, and J. Chan, "Equidensitometry by Coherent Optical Filtering," *Appl. Opt.* 15, 2394 (1976).
10. S. R. Dashiell and A. A. Sawchuck, "Nonlinear Optical Processing: Non-monotonic Halftone Cells and Phase Halftones," *Appl. Opt.* 16, 1936 (1977).
11. H. K. Liu and J. W. Goodman, "A New Coherent Optical Pseudo-Color Encoder," *Nouv. Rev. Optique*, 1, 285 (1976).
12. H. K. Liu, "Coherent Optical Analog-to-Digital Conversion Using a Single Halftone Photograph," *Appl. Opt.* 17, 2181 (1978).
13. F.T.S. Yu and A. Tai, "Incoherent Image Addition and Subtraction: A Technique," *Appl. Opt.* 18, 2705 (1979).
14. G. Indebetouw, "Production of Color Coded Equidensities Using Nonlinear Filtering, *Appl. Opt.* 16, 1951 (1977).
----, "New Method of Pseudo-Color Equidensitometry," *Appl. Opt.* 18, 4206 (1979).
15. B. H. Soffer, *et al.*, "Variable Grating Mode Liquid Crystal Device for Optical Processing," *SPIE* 218, 81 (1980).

16. W. Schneider, "Spatial Pulse Modulation for Coherent Optical Densitometry," *Opt. Acta.* 21, 263 (1974).
17. Ch. Thum., "Histograms of Continuous Tone Image by Transformation and Integration," *Angewandte Optik, Annual Report*, 31 (1980).
18. K. A. Stetson, *Optik* 22, 386 (1969).
19. A. A. Sawchuk, "Artificial Stereo," *Appl. Opt.* 17, 3869 (1978).
20. C. Warde, "Development and Applications of the Microchannel Spatial Light Modulator," *Interim Report, AFO SK-77-3328*, Nov. 28, 1980.
21. H. K. Liu, "Halftone Screen with Cell Matrix," U.S. Patent No. 4188225, Feb., 1980.
22. H. K. Liu and M. A. Karim, "Analysis of Image Reproduction Using a Linear Line Screen," *Opt. Lett.* 4, 408 (1979).
23. Tien-Hsin Chao, Hua-Kuang Liu, "Real-Time Optical Edge Enhancement using a Hughes Liquid Crystal Light Valve," *OSA Spatial Light Modulator Topical meeting* (1988)
24. C. Ward, H. Lamela-Rivera, H. K. Liu, "High-Gamma Spatial Light Modulator for Nonlinear Optical Processing," *SPIE* 613 (1986).
25. Hua-Kuang Liu, Jeffrey A. Davis, Roger A. Lilly, "Optical-Data-Processing Properties of a Liquid-Crystal Television Spatial Light Modulator," *Opt. Lett.* 10, 12 (1985).
26. Hua-Kuang Liu, Tien-Hsin Chao, "On the Progress of the Liquid Crystal Television Spatial Light Modulator," *OSA Spatial Light Modulator Topical Meeting* (1988).
27. Fai Mok, Joseph Diep, Hua-Kuang Liu, Demetri Psaltis, "Real-time Computer-Generated Hologram by Means of Liquid-Crystal Television Spatial Light Modulator," *Opt. Lett.* 11, 11 1986.
28. Tien-Hsin Chao, Hua-Kuang Liu, "Optical Pseudocolor Image Enhancement with Real-Time Large Screen Display," *Opt. Eng.* 27(5), (1988).
29. S. Y. Kung, H. K. Liu, "An Optical Inner-Product Array Processor for Associative Retrieval," *OE-LASE, Nonlinear Optics*, Los Angeles, CA, January 19-24 (1986).
30. Hua-Kuang Liu, S. Y. Kung, Jeffrey A. Davis, "Real-Time Optical Associative Retrieval Technique," *Opt. Eng.* 25(7), (1986).
31. Hua-Kuang Liu, Tien-Hsin Chao, Jacob Barhen, Graf Bittner, "Inner-Product Optical Neural Processing and Supervised Learning," *IEEE First Neural Networks Meeting*, Boston, MA, Sept. (1988).
32. Hua-Kuang Liu, Li-Jen Cheng, "Infrared Predetection Dynamic Range Compression via Photorefractive Crystal," *Appl. Opt.* 27(6), (1988).
33. Hua-Kuang Liu, Tien-Hsin Chao, "Optical Image Subtraction Techniques 1975-1985", *Hybrid Image Processing, SPIE 638*. (1986).

APPENDIX A

**High-Gamma Microchannel Spatial
Light Modulator**

**High-Gamma Microchannel Spatial Light
Modulator**

Semi-Annual Technical Report

February 1988

JPL Contract No. 182/381

Submitted by: C. Warde, MIT

Bistable Configurations of the MSLM

Suzanne Lau and Cardinal Warde

Massachusetts Institute of Technology

Department of Electrical Engineering and Computer Science

Cambridge, Massachusetts 02139

The performance of two MSLM configurations that can be used to achieve optically bistable operation are discussed. The operating criteria for each mode are investigated and the results presented.

1 Introduction

The development of a high resolution two-dimensional array of bistable optical circuits that offer low switching threshold intensity and high framing rate would be of significant benefit to the fields of digital optical processing, neural network processing, optical computer and VLSI interconnections, in addition to general purpose optical information processing. Such bistable operation could be used to perform 2-D logic operations, 2-D latches for digital memory, level restoration, level amplification, 2-D analog-to-digital conversion, thresholding, real-time image halftoning, sample and hold operations, and 2-D optically controlled switches.

To overcome the accuracy and dynamic range limitations of analog optical systems, digital optical systems have been proposed for a wide variety

of applications in numerical optical processing. These applications include such areas as real and complex number arithmetic and matrix algebra.

In this paper, we describe two possible Microchannel Spatial Light Modulator (MSLM) [1,2,3] configurations which can be used to obtain bistable optical operation: 1) the Fabry-Perot MSLM and 2) a closed-loop optical feedback configuration.

2 Microchannel Spatial Light Modulators

2.1 Standard MSLM

The standard Microchannel Spatial Light Modulator is a versatile, real-time image processing device which exhibits high optical sensitivity and a high framing speed. In its simplest form, it consists of a photocathode, microchannel plate (MCP), a planar acceleration grid and an electro-optic crystal plate (see Figure 1). The crystal has a high resistivity dielectric mirror on the side that faces the grid, and a conducting electrode on the other. The voltage sources V_{MCP} , V_{grid} and V_b are external to the device.

In the electron-deposition mode, the write beam (coherent or incoherent light) incident on the photocathode creates an electron image which is amplified by the MCP and proximity focused onto the dielectric mirror. The resulting spatially varying electric field modulates the refractive index of the crystal. Thus, the readout light which makes a double-pass through the crystal is phase or amplitude modulated, depending on the crystal cut and readout scheme (polarization or interferometric) employed.

The image is erased by flooding the photocathode with light so that the electrons are removed from the mirror by secondary electron emission. Alternatively, the device can be operated in the reverse mode, in which the image is written by removing charge from the dielectric mirror surface by secondary electron emission and erased by adding charge to the mirror. In the linear operating mode, the incremental surface charge density $\sigma(E)$ deposited on the crystal is proportional to the exposure E such that,

$$\sigma(E) = \left(\frac{\eta e G}{h\nu} \right) E \quad (1)$$

where η is the photocathode quantum efficiency, e is the electronic charge, G is the MCP gain, h is Planck's constant and ν is the frequency of the write light.

The high exposure sensitivity, high spatial resolution, long storage time and cascadable nature of the MSLM make it an especially versatile spatial light modulator. Optically-addressed MSLMs have already been utilized in adaptive phase compensation [4], optical image processing (such as contrast reversal, addition, subtraction, thresholding, edge enhancement and logic operations) [1]. More recently, an electron-beam addressed (e-beam) MSLM has been developed and used in matrix multiplication, with current research focused on implementing an optical inference machine for symbolic processing [5]. In the e-beam MSLM, the photocathode and the input light source are replaced by an electron gun whose beam is focused directly onto the MCP input face; images are written by controlling the electron beam using deflection grids.

The electro-optic crystal plate is the most important component de-

termining the MSLM performance. The mechanism for modulating the readout light is the linear electro-optic effect induced by an electric field between the spatially-varying charge density on the dielectric mirror and the transparent conducting electrode. For Pockels effect crystals, the induced phase change $\phi_{y'}, \phi_{x'}$ in the crystal is proportional to the surface charge density, $\sigma(E)$ (where x' and y' are the induced crystal axes). The exact expression depends on the type and cut of the crystal employed. For example, in oblique-cut LiNbO₃, [2] the electrically-induced phase retardation Γ is of the form,

$$\Gamma(E) = \phi_{y'} - \phi_{x'} = \frac{\pi\sigma(E)}{\lambda C} (n_e^3 r_{y'} - n_o^3 r_{x'}) \quad (2)$$

where λ is the wavelength of the readout light, C is the average capacitance per unit area of the crystal, n_o and n_e are the ordinary and extraordinary refractive indices respectively, and $r_{x'}$ and $r_{y'}$ are effective electro-optic coefficients. The amount of charge density $\sigma(E)$, or equivalently the voltage across the crystal, V_s , uniquely determines the amount of phase modulation. When such a crystal is read out between crossed polarizers, the transmittance of the crystal is given by,

$$T = \frac{I_r}{I_i} = \sin^2 \left(\frac{\Gamma(E)}{2} \right) \quad (3)$$

where the phase retardation Γ is a function of input exposure E .

2.2 Fabry-Perot MSLM

The Fabry-Perot version of the MSLM (F-P MSLM)[6] employs a crystal that functions as an electro-optically tunable, Fabry-Perot etalon when

electrons are deposited or removed from its surface. To fabricate such a crystal, standard dielectric mirrors are deposited on both surfaces of the crystal and then a transparent conducting overcoat is deposited on the readout surface of the crystal.

For a z-cut LiNbO₃ device operated in the reflex mode, the phase retardation Γ which results from the electrically induced refractive index change Δn , is, to a good approximation, proportional to the surface charge density $\sigma(E)$ and to the modulator write light exposure E . For this crystal cut, Γ is given by

$$\Gamma(E) = \phi_z = \phi_y = \frac{2\pi n_o^3 r_{13} \sigma(E)}{\lambda C} \quad (4)$$

where ϕ_z and ϕ_y are the phase changes along the respective crystal axes, λ is the wavelength of the readout light and V_z is the voltage across the crystal.

For a Fabry-Perot etalon with surfaces of reflectivity R , it is well known that the ratio T , of the total reflected intensity to the incident intensity, is given by,

$$T = \frac{I_r}{I_i} = \frac{4R \sin^2\left(\frac{\pi}{2}\right)}{(1 - R)^2 + 4R \sin^2\left(\frac{\pi}{2}\right)} \quad (5)$$

This equation can be rewritten as,

$$T = \frac{I_r}{I_i} = \frac{F \sin^2\left(\frac{\pi}{2}\right)}{1 + F \sin^2\left(\frac{\pi}{2}\right)} \quad (6)$$

where the coefficient of finesse F is defined as,

$$F \equiv \frac{4R}{(1 - R)^2} \quad (7)$$

Figure 2 is a plot of T vs. ϕ ; the voltage across the crystal, V_z , corresponding to the charge density $\sigma(E)$, uniquely determines the phase modulation (see Equation 4). Note that the reflected readout transmission of the F-P MSLM approaches zero when ϕ takes on integer multiples of 2π radians. The halfwave voltage V_π is defined as the voltage required for phase modulation of π radians. For z-cut LiNbO₃, the calculated V_π is 3100V in reflection. An important measure of the sharpness of the Fabry-Perot fringes is the finesse \mathcal{F} , which is defined as the ratio of the separation of adjacent maxima to the full width at half maximum ($FWHM$) and is given by,

$$\mathcal{F} = \frac{\pi\sqrt{F}}{2} = \frac{\pi\sqrt{R}}{(1-R)} \quad (8)$$

Thus, a Fabry-Perot MSLM with high \mathcal{F} will have rapid rise in intensity on either side of a zero. The \mathcal{F} can be expressed in terms of ϕ , V_z or the induced pathlength Δ through the etalon, ($\phi = \frac{2\pi n_e}{\lambda} \Delta$),

$$\mathcal{F} = \frac{2\pi}{FWHM(\phi)} = \frac{2V_\pi}{FWHM(V_z)} = \frac{\lambda}{FWHM(\Delta)} \quad (9)$$

where λ is the wavelength of the readout light and $FWHM(\cdot)$ is the full width at half maximum expressed in terms of ϕ , V_z , or Δ (see Figure 2).

The reflectivity R of the dielectric mirrors deposited on the crystal surfaces and the plane-parallelism of the crystal surfaces are the determining factors of the F-P MSLM performance. R determines the finesse \mathcal{F} of the etalon (Equation 8). By definition, the higher the \mathcal{F} , the narrower the $FWHM$ and the more sensitive the etalon becomes to deviations of the crystal surfaces from plane-parallelism.

3 F-P MSLM Image Processing Criteria

We fabricated a F-P MSLM that had a crystal coated with mirrors of $R = 0.7$. This corresponds to a finesse $\mathcal{F} \approx 8.76$. The crystal had a single Haidinger fringe around its 25mm perimeter before device assembly; this corresponds to crystal surface plane-parallelism of about $\approx \pm\lambda/4$. No 2-D imaging was attempted using this experimental system, since the crystal did not have a large enough area of uniform modulation.

The physical limitation on the finesse \mathcal{F} of a F-P MSLM that would be useful in image processing, is determined by the deviation of the polished electro-optic crystal plate surfaces from plane-parallelism. By definition, the $FWHM(\Delta)$, expressed in terms of optical pathlength in the etalon, is a measure of the tolerance of the etalon to deviations from plane-parallelism of the crystal faces. That is, a region of the crystal with thickness l_z and all other areas of the crystal with thickness l , will modulate together if l satisfies the following,

$$l_z - \frac{FWHM(\Delta)}{2} \leq l \leq l_z + \frac{FWHM(\Delta)}{2} \quad (10)$$

This can be rewritten,

$$|l - l_z| \leq \frac{FWHM(\Delta)}{2} \quad (11)$$

Let us define a measure of the crystal plane-parallelism a as,

$$a \equiv |l - \bar{l}_z| \quad (12)$$

where \bar{l}_z is the average crystal thickness, and l is the crystal thickness at any point of the crystal. To attain uniform modulation of the Fabry-

Perot, the following condition (derived from Equation 11) must be satisfied throughout the modulating region,

$$a \leq \frac{FWHM(\Delta)}{2} = \frac{\lambda}{2\mathcal{F}} \quad (13)$$

That is,

$$\mathcal{F} \leq \frac{\lambda}{2a} \quad (14)$$

The result of violating this condition was aptly demonstrated by the prototype F-P MSLM built, where $a \approx \lambda/4$ and $\mathcal{F} \approx 8.76$. Substituting into Equation 14,

$$\mathcal{F} = 8.76 \not\leq \frac{\lambda}{2\lambda/4} = 2 \quad (15)$$

Alternatively stated, $\mathcal{F} = 8.76$ ($R = 0.7$) would require a crystal with $a \leq \lambda/17.5$. Thus, the performance of high \mathcal{F} F-P MSLMs will be severely limited by the maximum achievable plane-parallelism of the polished crystal surfaces by present technology. With the present technology capable of achieving a maximum $a \approx \lambda/4$, \mathcal{F} is restricted to $\mathcal{F} \leq 2$ (corresponding to $R \leq 0.25$). These limitations are not easily overcome using state-of-the-art polishing techniques, especially for the thin crystals required.

A possible alternative would be the use of a thin film modulator material deposited onto an optically flat glass substrate. This would simultaneously ensure the plane-parallelism of the modulator and improve the spatial resolution of the modulator.

A method for circumventing the plane-parallelism requirement (and enable use of the presently available crystals), has been suggested for use with visible-photocathode F-P MSLMs [3]. In this case, the crystal par-

allelism deviations are initially compensated with the appropriate charge distribution before any image processing is done. This is accomplished by an optical feedback configuration where the reflection of a uniform readout light is fed back to the photocathode until the internal device nonuniformities are compensated and a stable equilibrium has been attained. This method is discussed at length below.

4 Optical Feedback Configuration

The optical feedback configuration of Figure 3 is a means for: (1) compensating phase variations introduced by the modulator crystal, and (2) realizing bistable operation. Phase compensation is especially important for high finesse Fabry-Perot MSLMs which would otherwise require crystals with extreme plane-parallelism unattainable by present polishing technology.

In Figure 3, the system or input object under investigation is placed in the input plane (at P_1) or in the feedback path (at P_2). A variable spatially uniform amplitude attenuation plate is also placed in the feedback path at P_2 , to control the feedback level. The output is recorded in plane P_3 . Note that in this case the intensity of the signal fed back is proportional to the intensity of the output signal beam.

The crossed polarizers in Figure 3 are omitted for z-cut crystals since these are read out interferometrically. The MSLM may be operated in the closed loop (feedback) mode by opening the shutter S, and in the open loop mode by closing the shutter. The MSLM output is monitored in the

plane P_3 and is equal to some fraction of the feedback signal. A procedure for operating both a standard or Fabry-Perot MSLM in this configuration would be 1) compensate, closed loop 2) bias crystal, open loop 3) write nonlinear image, closed loop. Phase compensation and bistable operation are accomplished as follows. The input object is removed (uniform input), the feedback path is closed (S open) and the device is operated, preferably in the electron deposition mode, until the readout intensity in plane P_3 is uniformly zero.

In this closed loop operation, areas of non-zero transmittance of the crystal will result in feedback to the photocathode and deposition of surface charge on the corresponding areas of the crystal so as to drive the transmittance of the modulator at that region to zero. The intrinsic phase variations of the MSLM are thus compensated with the appropriate amount of surface charge and a stable equilibrium is attained.

The device should now be biased to a bright background by operation in the open loop mode (S closed) flooding the photocathode with light and adjusting V_b to add or remove a uniform charge density while still maintaining the phase compensation charge profile. The flood light is then shut off and V_b set to a large enough voltage such that the gap voltage V_g remains high enough to ensure linear operation throughout the imaging process.

The image transparency to be processed is then placed in the input plane P_1 and transferred by lens L_0 onto the crystal with the shutter S open. The resulting feedback signal is once again reimaged via the feedback path

onto the photocathode. In this case, for a given fixed incident intensity on the input image, the high transmittance regions of the input image will at first result in strong feedback and rapid deposition of sufficient surface charge on the crystal so as to drive the modulator transmittance of those points to zero (stable equilibrium) within the allowed feedback time t_w . On the other hand, low transmittance regions of the input image result in weak feedback, which will in turn result in very small amounts of additional charge deposition on the crystal; thus, at the end of the feedback time t_w , the transmittance of the modulator remains high. The MSLM stores these charge patterns and when read out in open loop mode using a uniform source, a bistable image of the input transparency results.

The theory describing the MSLM operation in this optical feedback configuration can best be modelled by considering the behaviour of a single pixel of the modulator and neglecting any transverse effects from neighboring pixels.

The intensity of the feedback signal, $I_f(t)$ can be written as,

$$I_f(t) = I_i T(t) \quad (16)$$

where I_i is the incident input intensity transferred to the crystal, $T(t)$ is the particular time-dependent transmittance and t is the elapsed feedback time. $I_o(t)$, the output signal is some fraction of the feedback signal (see Figure 4). For a standard oblique-cut MSLM readout between crossed polarizers, the transmittance $T(t)$ (from Equations 2 and 3) is given by,

$$T(t) = \sin^2 \left(\frac{\Gamma(t)}{2} \right) = \sin^2 \left[\frac{\pi \sigma(t)}{2\lambda C} (n_i^3 r_y - n_o^3 r_x) \right] \quad (17)$$

where $\sigma(t)$ is the total accumulated surface charge density on the crystal at time t . In the linear mode of operation the incremental charge accumulated in the time dt is derived from Equation 1 and is given by

$$d\sigma(t) = \left(\frac{\eta e G}{h\nu} \right) I_f(t) dt \quad (18)$$

The total charge density integrated on the surface of the crystal at time t is therefore,

$$\sigma(t) = \int_0^t \left(\frac{\eta e G}{h\nu} \right) I_f(t') dt' + \sigma_0 \quad (19)$$

where σ_0 is the initial charge on the crystal. Equation 16 can now be rewritten as,

$$I_f(t) = I_i \sin^2 \left[B \int_0^t I_f(t') dt' + \frac{\phi_0}{2} \right] \quad (20)$$

where ϕ_0 is the phase retardation due to σ_0 and B is a lumped MSLM constant. For the purposes of this analysis, the lumped constant B , which acts simply as a scaling parameter, is assumed to be equal to 1. The initial conditions at $t = 0$ is,

$$I_f(t = 0) = I_i \sin^2 \left(\frac{\phi_0}{2} \right) \quad (21)$$

which is the reflection due to the initial bias of the crystal. Notice that at time t_f ,

$$I_f(t_f) = I_i \sin^2 \left[\int_0^{t_f} I_f(t') dt' + \frac{\phi_0}{2} \right] = 0 \quad (22)$$

if

$$\int_0^{t_f} I_f(t') dt' + \frac{\phi_0}{2} = p\pi; \text{ where } p = 0, \pm 1, \pm 2, \dots \quad (23)$$

and moreover,

$$0 \leq \int_0^{t_f} I_f(t') dt' = p\pi - \frac{\phi_0}{2} \leq \pi \quad (24)$$

Once the feedback signal $I_f(t)$ is driven to zero at time t_f , no further charge deposition occurs, so that,

$$I_f(t) = 0 \text{ for all } t \geq t_f \quad (25)$$

and stable equilibrium is achieved. Notice that for phase compensation, the shutter S must remain open long enough to allow the entire crystal to achieve equilibrium. This behavior is monitored in the output plane.

Alternatively, for nonlinear processing (e.g. bistable operation) feedback is allowed only long enough for areas of the input with transmittance above some threshold to be driven to equilibrium. The threshold transmittance is determined by the feedback time t_w . In the nonlinear processing case, t_w is such that,

$$\int_0^{t_w} I_f(t') dt' + \frac{\phi_0}{2} \approx \begin{cases} p\pi & \text{for high input transmittance} \\ \frac{\phi_0}{2} & \text{for low input transmittance} \end{cases} \quad (26)$$

Thus the modulator transmittance will be given by,

$$T(t_w) \approx \begin{cases} 0 & \text{for high input transmittance} \\ \sin^2\left(\frac{\phi_0}{2}\right) & \text{for low input transmittance} \end{cases} \quad (27)$$

and the desired bistable image will be attained upon readout in the open loop configuration with a uniform source.

The dynamic behaviour of the feedback signal is best analysed by numerically calculating and plotting $I_f(t)$. From the fundamental theorem of calculus, Equation 20 can be estimated as follows,

$$I_f(t = N \Delta t) = I_i \sin^2 \left[\sum_{n=1}^N I_f((n-1), \Delta t) \Delta t + \frac{\phi_0}{2} \right] \quad (28)$$

where the feedback signal is considered constant during the incremental time interval $\Delta t = t/N$; N is the number of time intervals. The initial condition is still given by Equation 21.

Equation 28 can be analysed for two cases: (1) varying the initial bias phase ϕ_0 due to the crystal nonuniformities, and (2) varying the input intensity I_i transmitted by the input transparency.

Figure 5a contains plots of $I_f(t)$ for multiple values of the initial phase ϕ_0 with constant input intensity I_i . The transmission will decay to zero faster than the characteristic sine-squared curve and equilibrate. Notice that in cases where $\phi_0/2$ occurs before a maximum of modulator transmission, the feedback signal will actually increase first before decaying to zero.

The effect of different input intensities for the same initial phase $\phi_0 = \pi$ on $I_f(t)$ is shown in Figure 5b, and the effect on the modulator transmittance $T(t)$ is shown in Figure 5c. In this case, ϕ_0 is set at π radians and the relative values of the input intensities varied. Notice that $I_f(t)$ (and thus, $T(t)$) decays to zero more rapidly for higher input intensity. This supports the discussion on the use of this feedback configuration for nonlinear applications. At some time t_w , indicated in Figure 5c regions with high I_i are driven to zero transmittance, while those with low I_i leading to very little additional charge deposition, maintain relatively high modulator transmittance.

A similar analysis has been done for the dynamic behaviour of the Fabry-Perot MSLM. For this version of the device, the transmittance $T(t)$ is

derived from Equations 5 and 16 rewritten as

$$I_f(t) = I_i \frac{4R \sin^2 \left[B' \int_0^t I_f(t') dt' + \frac{\phi_0}{2} \right]}{(1 - R)^2 + 4R \sin^2 \left[B' \int_0^t I_f(t') dt' + \frac{\phi_0}{2} \right]} \quad (29)$$

where ϕ_0 is again the initial bias phase and B' is a lumped device constant. Once again, B' is a scaling parameter and is assumed equal to 1. The initial condition is,

$$I_f(0) = \frac{I_i 4R \sin^2 \left(\frac{\phi_0}{2} \right)}{(1 - R)^2 + 4R \sin^2 \left(\frac{\phi_0}{2} \right)}. \quad (30)$$

The feedback signal will again equilibrate at zero when $t \geq t_f$, where t_f is still as defined in Equation 23. The Fabry-Perot device will operate in much the same manner as the standard device.

The high finesse Fabry-Perot device will compensate for crystal deviations from plane-parallelism if sufficient feedback time is allowed, and can achieve nonlinear operation when the feedback time is appropriately limited. Equation 29 was analysed numerically and plots were made for: 1) $I_f(t)$ with constant I_i , and different ϕ_0 (shown in Figure 6a), and 2) $I_f(t)$ and the corresponding $T(t)$ with constant $\phi_0 = \pi$ and different relative I_i (shown in Figure 6b and c).

Notice that the intrinsic nonlinear characteristics of a high finesse Fabry-Perot crystal ($R = 0.7$) result in similar nonlinear behaviour of the feedback signal and the modulator transmittance. In nonlinear operation, the threshold input transmittance can be precisely determined by t_w .

The disadvantage of using this type of feedback configuration to compensate for crystal deviations from plane-parallelism by charge deposition, is that there is a trade-off in dynamic operating range. Experimentally the

safe dynamic operating range of an MSLM device is about 4kV; this limitation is imposed by arcing between components and stresses due to high electrostatic potentials. Consider a standard device with $V_{rR} = 1250V$; if the MSLM to be compensated has a crystal with plane-parallelism factor (see Equation 12), $a = \lambda/4$, then the crystal will require about $V_{rR}/2 = 625V$ equivalent charge compensation in some areas. Thus, about 15% of the available operation range is lost for compensation, leaving over 3kV with which to work; this is not unreasonable for a crystal with $V_{rR} = 1250$ (full modulation can be realised for a grid voltage of 1 – 2kV).

However, the compensation trade-off is much more severe in the case of the Fabry-Perot. Consider a device with a z-cut LiNbO₃ crystal, $V_r = 3100V$; once again assuming $a = \lambda/4$, the crystal will require $V_r/2 = 1550V$ equivalent charge compensation. This is about 40% of the available operating range, leaving only about 2450V with which to work. Although this range is greater than the voltage required for a light to dark transition (600 – 800V; $R = 0.7$), if the grid voltage is 1 – 2kV, no allowance would be made for the fact that the thickness of the crystal could be close to $(n + \frac{1}{2})\lambda$, where n is an integer and λ is the wavelength of the readout light. Thus, the crystal would be biased in the neighborhood of the central flat region of the Fabry-Perot characteristic. This additional fraction of a wavelength must be accounted for experimentally by properly biasing the operating voltages and in the worst case can require as much as $V_r = 3100V$.

The ability to compensate internal device phase variations is critical for imaging with a Fabry-Perot MSLM using currently available crystals.

However, the amount of phase compensation that can be tolerated while leaving sufficient operating range, is limited by the precise crystal thickness deviation from an integer multiple of λ , and the voltage limitations imposed on the device.

5 Summary

The MSLM can be used to implement optical bistable circuits in either its Fabry-Perot configuration, or in an optical feedback mode. The optical feedback mode may also be used to compensate plane-parallelism deviations of the electro-optic crystal. The threshold level is established by the initial bias before operation of the MSLM. The relative simplicity of these all-optical architectures are particularly attractive in more sophisticated digital optical processing applications.

Acknowledgements

This work was supported in part by a subcontract with the Jet Propulsion Laboratory, California Institute of Technology, under prime contract No. NASA/JPL Contract NAS7-918, Task Order RF-182, Amendment No. 381, with the Army Research Office.

References

- [1] C. Warde, A. M. Weiss, A. D. Fisher and J. I. Thackara, "Optical information processing characteristics of the microchannel spatial light modulator," *Applied Optics*, 20, 2066, 1981.
- [2] C. Warde and J. I. Thackara, "Oblique-cut LiNbO₃ microchannel spatial light modulator," *Optics Letters*, 7, 344, 1982.
- [3] S. D. Lau, "Variable-Gamma Spatial Light Modulator," *SM Thesis*, Massachusetts Institute of Technology, June 1987.
- [4] A. D. Fisher, "Techniques and Devices for High-Resolution Adaptive Optics," *PhD Thesis*, Massachusetts Institute of Technology, August 1981.
- [5] J. A. Kottas, "Optical Pattern Processors for Matrix Multiplication and Symbolic Inference," *SM Thesis*, Massachusetts Institute of Technology, February 1986.
- [6] C. Warde, H. Lamela-Rivera and H. K. Liu, "High-Gamma spatial light modulator for nonlinear optical processing," in *Proc. SPIE Nonlinear Optics and Applications*, 613, 153, 1986.

FIGURE CAPTIONS

Figure 1: Microchannel Spatial Light Modulator

Figure 2: Readout characteristic for a Fabry-Perot MSLM

Figure 3: Optical feedback configuration of the MSLM

Figure 4: Functional block diagram of the closed-loop configuration

Figure 5: Output intensity $I_f(t)$ of a standard MSLM in the closed-loop configuration: a) for constant input intensity I_i with initial phase ϕ_0 as a parameter, b) for constant initial phase $\phi_0 = \pi$ and with input intensity I_i as a parameter, and c) corresponding modulator transmission $T(t)$

Figure 6: Output intensity $I_f(t)$ of a Fabry-Perot MSLM in the closed-loop configuration: a) for constant write intensity I_i with initial phase ϕ_0 as a parameter, b) for constant initial phase $\phi_0 = \pi$ and with write intensity I_i as a parameter, and c) corresponding modulator transmission $T(t)$

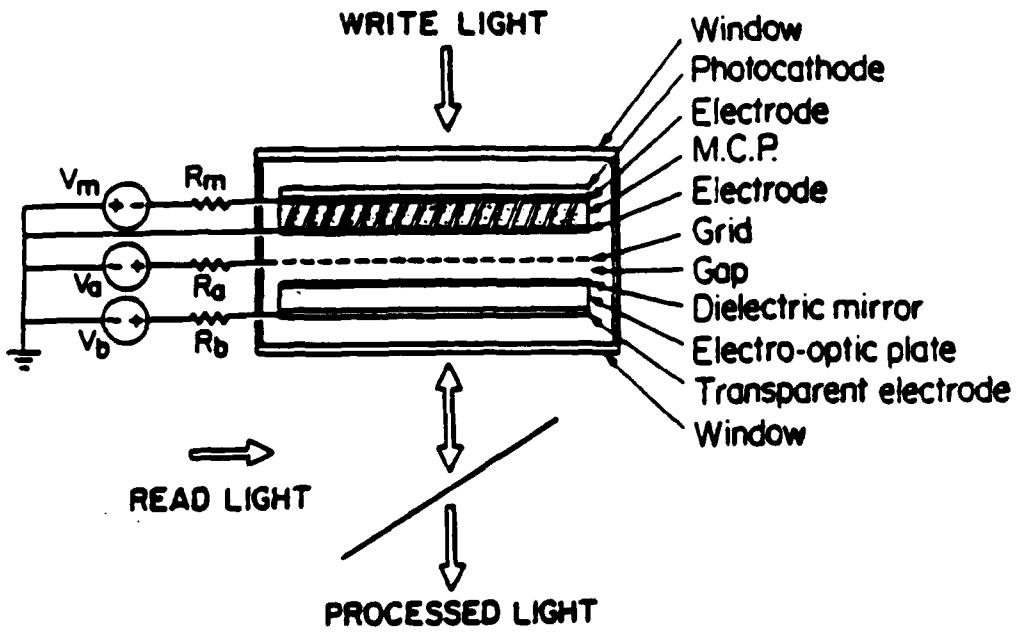


Figure 1: Microchannel Spatial Light Modulator

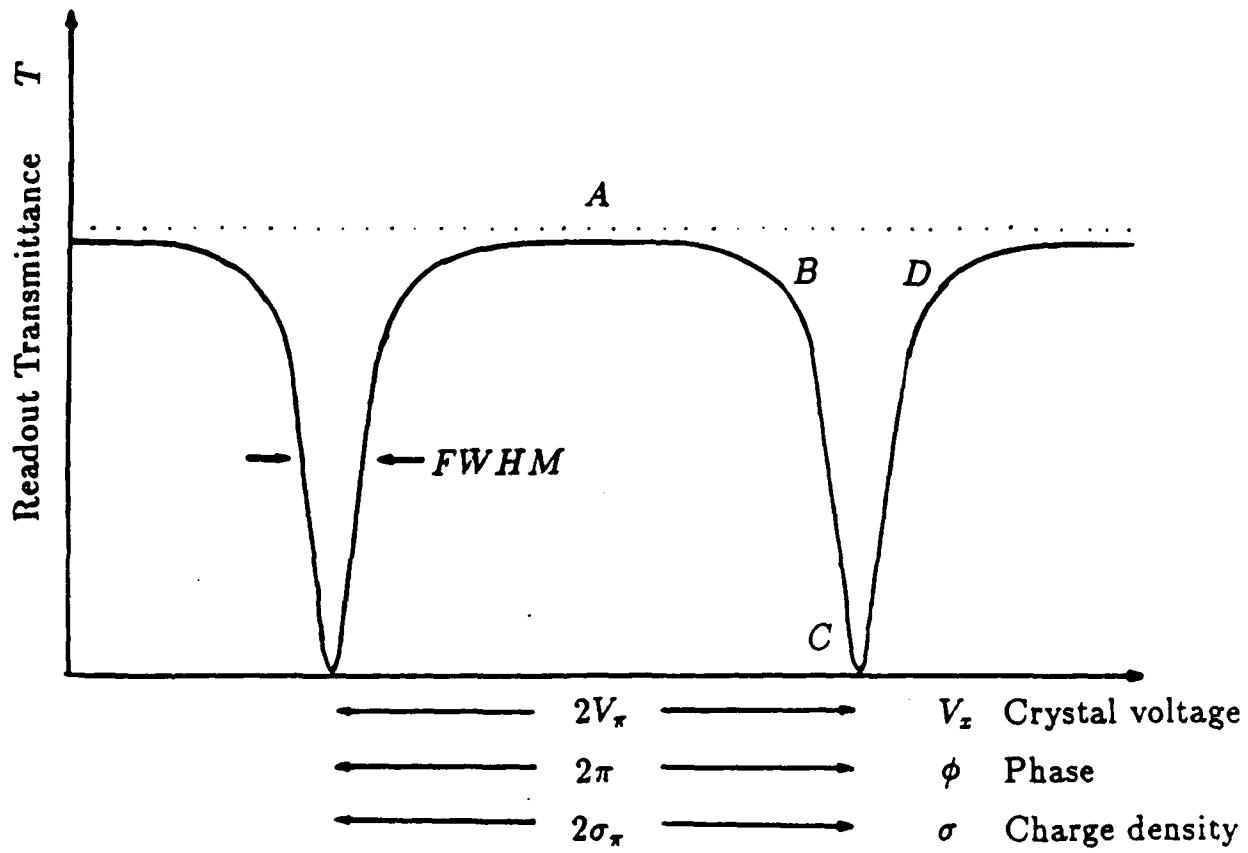


Figure 2: Readout characteristic for a Fabry-Perot MSLM

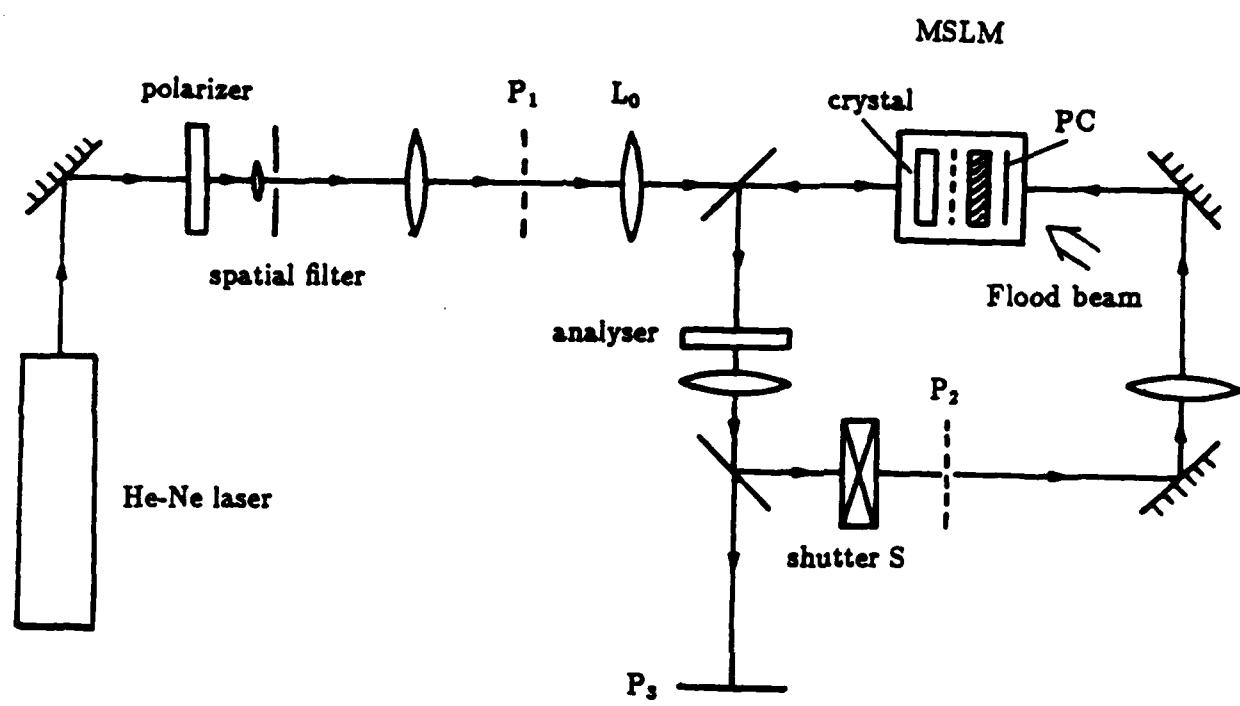


Figure 3: Optical feedback configuration of the MSLM

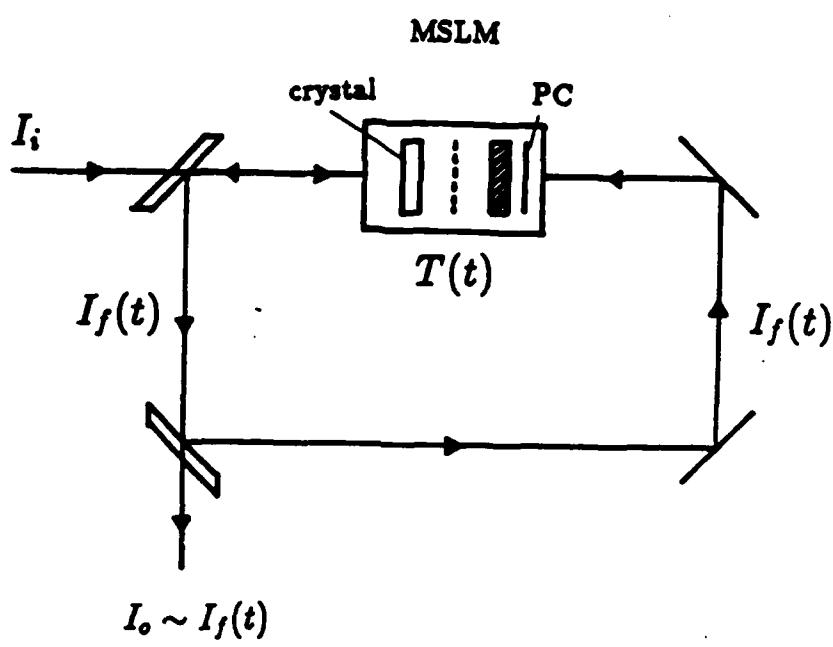


Figure 4: Functional block diagram of the closed-loop configuration

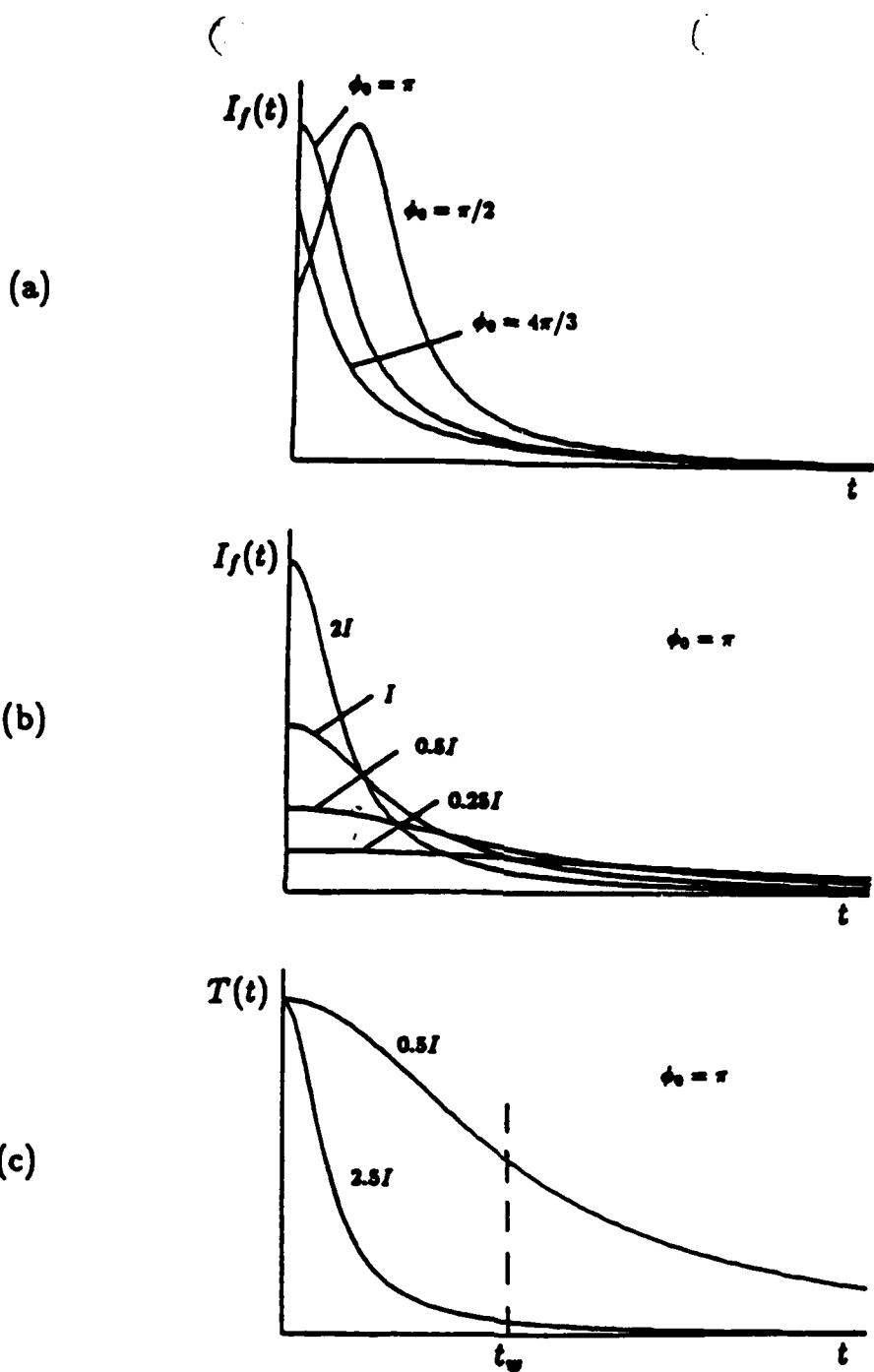


Figure 5: Output intensity $I_f(t)$ of a standard MSLM in the closed-loop configuration: a) for constant input intensity I_i with initial phase ϕ_0 as a parameter, b) for constant initial phase $\phi_0 = \pi$ and with input intensity I_i as a parameter, and c) corresponding modulator transmission $T(t)$

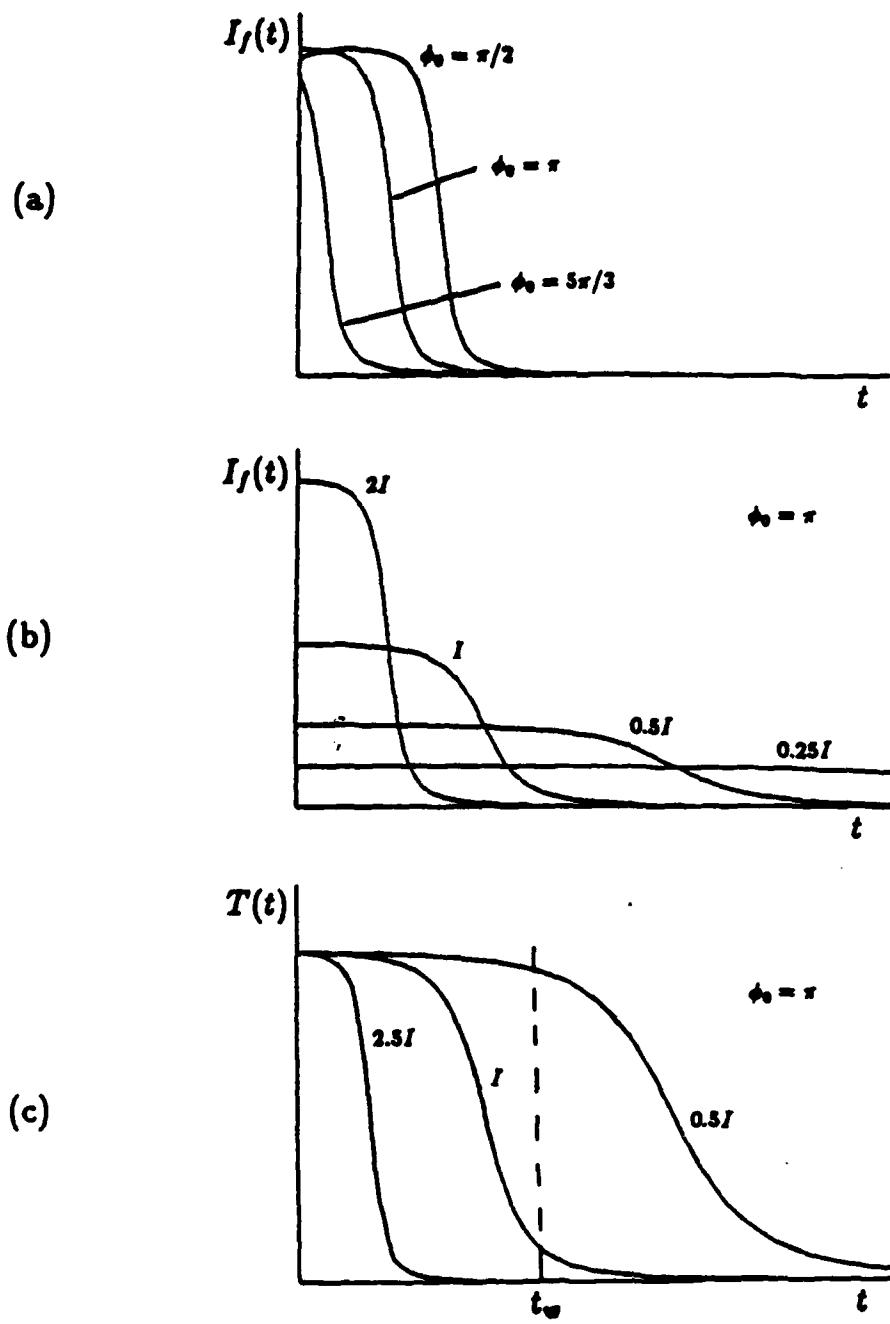


Figure 6: Output intensity $I_f(t)$ of a Fabry-Perot MSLM in the closed-loop configuration: a) for constant write intensity I_i with initial phase ϕ_0 as a parameter, b) for constant initial phase $\phi_0 = \pi$ and with write intensity I_i as a parameter, and c) corresponding modulator transmission $T(t)$

APPENDIX B

**Variable-Gamma Spatial
Light Modulator**

**VARIABLE-GAMMA
SPATIAL LIGHT MODULATOR**

by

Suzanne Dorothy Lau

**B.E.S., Johns Hopkins University
(1984)**

Submitted in partial fulfillment
of the requirements for the degree of

**Master of Science
in Electrical Engineering and Computer Science**

at the

Massachusetts Institute of Technology

June 1987

© Massachusetts Institute of Technology, 1987

Signature of Author

Suzanne Lau
Department of Electrical Engineering and Computer Science
May 15, 1987

Certified by

Cardinal Warde

**Cardinal Warde
Thesis Advisor**

Accepted by

Arthur C. Smith
Chairman, Department Committee on Graduate Students

Acknowledgements

I wish to express my deepest gratitude to Cardinal Warde who has always been more than just a thesis advisor. His support and encouragement throughout the ups and downs of this thesis work have been greatly appreciated. I am especially grateful for his willingness to take time in the middle of a hectic afternoon or at the end of a busy day to discuss a concept or an idea.

I have been privileged to work with a close-knit and diverse group of colleagues. I wish to thank them all for making research fun. Especially: Jim for many insightful discussions, box-making skills, photographic expertise, borrowed equipment and calming presence in times of arcing; Scott and Doyle for teaching me about crystals, giving general advice, coffee and cookies; Ala for donating his likeness in the interest of research; and Kuang Yi for taking care of us all. Thanks also to Ling Yi for being around and getting me interested in this work.

Above all, I am indebted to my parents for their unwavering support of all my endeavours, both academic and otherwise. They have provided the solid foundation from which I have derived the confidence and ambition to overcome any obstacles in the pursuit of excellence. I am grateful to David, Jennifer and Stanley for making sure that I kept my feet on the ground and moving forward.

Finally, I wish to thank Thomas for everything and so much more.

The experimental work described in Chapter 6 was performed at M.I.T. in collaboration with H. K. Liu of the Jet Propulsion Lab. The author wishes to thank Hamamatsu Photonics K. K. and Ming Wu of the Hamamatsu Corporation of Japan for the loan of the vacuum-sealed MSLM used herein; and Richard Hertel of ITT Electro-Optical Products Division for his unlimited and speedy supply of acceleration grids.

Some of the results described in this thesis were presented at the Optical Society of America *Topical Meeting on Optical Computing* in Lake Tahoe, Nevada in March, 1987.

**VARIABLE-GAMMA
SPATIAL LIGHT MODULATOR**
by
Suzanne Dorothy Lau

Submitted to the
Department of Electrical Engineering and Computer Science
on May 15, 1987 in partial fulfillment of the requirements
for the degree of Master of Science.

Abstract

The lack of a general purpose, real-time, reusable spatial light modulator has long precluded full exploitation of the inherent speed and parallel processing capability of optical systems. Such a high-resolution, optically-addressed, variable-gamma device would be applicable to both linear (low gamma) and nonlinear (high gamma) optical processing. This thesis investigates the following techniques that are used to manipulate the switching characteristic of a Microchannel Spatial Light Modulator (MSLM): 1) converting the standard device to a Fabry-Perot MSLM, 2) operation in a closed-loop feedback configuration, 3) operation in the electron deposition/grid stabilized mode, 4) operation in the real-time hardclipped thresholding mode. The desired switching characteristic described herein is more suitable for characterizing the behaviour of this spatial light modulator than the conventional gamma defined for photographic film. A standard MSLM operated in the real-time hardclipped thresholding mode is used to demonstrate the application of a high-gamma MSLM in real-time halftone screen processing. Device improvements and recommendations for future work are also included.

Thesis Advisor: Cardinal Warde

Title: Associate Professor of Electrical Engineering and Computer Science

Contents

1	Introduction	11
2	Standard MSLM	15
2.1	Device Description	15
2.2	Principles of Operation	18
2.3	Linear Operating Modes	21
2.4	Experimental System	22
2.4.1	The Hamamatsu MSLM	22
2.4.2	Power Requirements	23
2.4.3	Optical System	26
2.4.4	Output Devices	29
2.5	Device Performance	30
3	Fabry-Perot MSLM	37
3.1	Device Description	37
3.2	Linear Operating Modes	40
3.3	Experimental System	41
3.3.1	Prototype Fabry-Perot MSLM	41
3.3.2	Power Requirements	44

3.3.3 Optical Systems	46
3.3.4 Output Devices	46
3.4 Device Performance	49
3.5 F-P MSLM Image Processing Criteria	51
4 Optical Feedback Configuration	55
5 Variable Gamma Operation	67
5.1 Switching Characteristics	68
5.1.1 Conventional Gamma Characteristic of Photographic Film	68
5.1.2 Desired Switching Characteristic of an MSLM	69
5.2 Electron deposition/Grid stabilized Mode	73
5.3 Real-Time Hardclipped Thresholding Mode	78
6 Optical Information Processing	85
6.1 Applications Discussion	85
6.2 Halftone Screen Process	87
7 Conclusion	97
7.1 Summary	97
7.2 Future Work	99
A Fabry-Perot MSLM Assembly	102
B Anomalous Observations	107
B.1 Fabry-Perot MSLM	107
B.1.1 Marked Decrease in Device Speed	107
B.1.2 Bias Voltage Drifting	108

B.2 Hamamatsu Standard MSLM	108
B.2.1 Unusually high V_{eR}	108
B.2.2 Nonuniform charge build-up	108
B.2.3 Dark Ring	109
B.2.4 Ghost Image	109
B.2.5 Image Curvature	109
B.2.6 Crystal Cracking	111

List of Figures

1.1 Microchannel Spatial Light Modulator	12
2.1 MSLM internal configuration	19
2.2 Hamamatsu MSLM	23
2.3 MSLM controller timing waveform	25
2.4 Standard MSLM biasing box	27
2.5 Optical configuration for Hamamatsu MSLM	28
2.6 Readout characteristic for a standard MSLM	32
2.7 Resolution chart image	36
3.1 Readout characteristic for a Fabry-Perot MSLM	39
3.2 Prototype Fabry-Perot MSLM	42
3.3 Fabry-Perot MSLM biasing box	45
3.4 Optical configuration for Fabry-Perot MSLM	47
3.5 Detector amplifier circuit	48
3.6 Fabry-Perot output image	50
3.7 Experimentally measured F-P MSLM readout characteristic	52
4.1 Optical feedback configuration of the MSLM	56
4.2 Functional block diagram of the closed-loop configuration	59

4.3	Output intensity $I_f(t)$ of a standard MSLM in the closed-loop configuration	62
4.4	Output intensity $I_f(t)$ of a Fabry-Perot MSLM in the closed-loop configuration	65
5.1	Conventional gamma characteristic for photographic film	68
5.2	MSLM Readout transmission characteristics	70
5.3	Desired MSLM switching characteristic	71
5.4	I_o vs. t_w for standard Hamamatsu MSLM	76
5.5	I_o vs. t_w for prototype Fabry-Perot MSLM	77
5.6	Harclipped thresholding with I_o as a parameter	80
5.7	Harclipped thresholding with t_w as a parameter	81
5.8	Hardclipped thresholding of the gray scale image of an M.I.T. student	84
6.1	Halftone screen processing: basic concept	88
6.2	Profile of the contact screen utilized	90
6.3	Optical System for implementing halftone processing	93
6.4	Real-time halftone screen processing results	95
6.5	More halftone screen processing results	96
A.1	Fabry-Perot flange specifications	104
B.1	Dark ring	110
B.2	Ghost image	110
B.3	Image curvature	112
B.4	Cracking phenomenon	114
B.5	Charge-induced poling explanation of oblique-cut LiNbO ₃	116

List of Tables

2.1 MSLM power requirements	24
3.1 Fabry-Perot MSLM components	43
3.2 Fabry-Perot MSLM power requirements	46
6.1 Pulse frequency modulation: Grating carrier frequency corresponding to I_{th}	91
A.1 Relevant dimensions of the Fabry-Perot MSLM components	103

Chapter 1

Introduction

The lack of a general purpose, real-time reusable spatial light modulator has long precluded full exploitation of the inherent speed and parallel processing capability of optical systems. Such a device is necessary as: 1) an input data plane processor, 2) an optical filter/interface in the Fourier plane and 3) a recording/storage material. Silver halide photographic film, currently the most commonly used optical processing component, possesses one major weakness; its wet development process is cumbersome and more importantly, cannot be carried out in real time.

A high-resolution, optically-addressed, variable-gamma spatial light modulator would be very useful for general purpose, real-time optical processing. This thesis describes how the standard and the Fabry-Perot Microchannel Spatial Light Modulators (MSLM) can be modified and operated to achieve a wide range of gamma analogous to that of photographic emulsions. Such a variable-gamma device will be applicable to both linear (low gamma) and nonlinear (high gamma) optical processing [1,2,3,4].

The standard Microchannel Spatial Light Modulator is a versatile, real-time image processing device which exhibits high optical sensitivity and a high framing

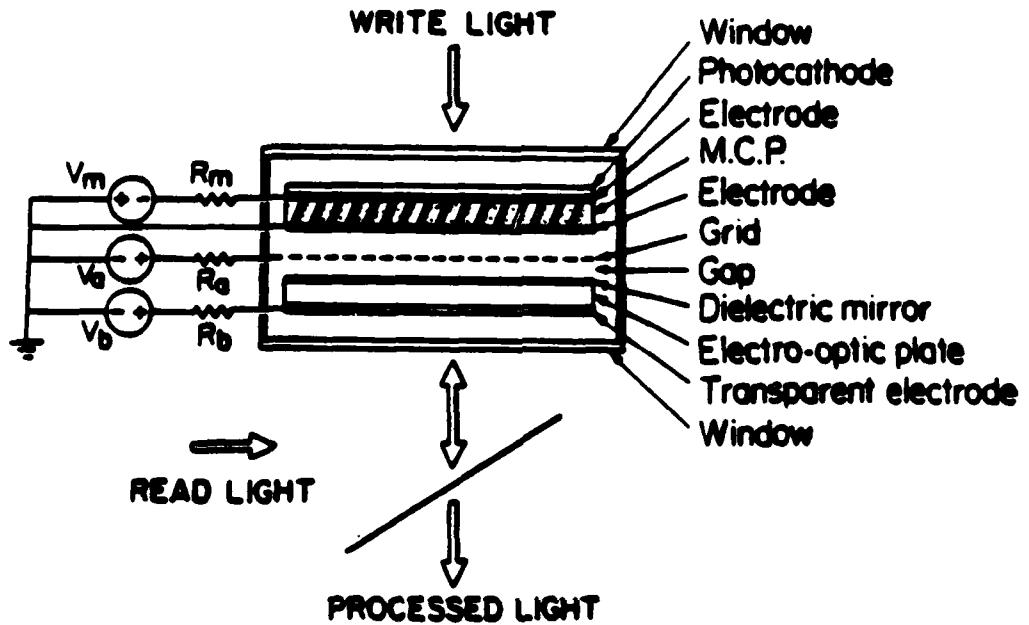


Figure 1.1: Microchannel Spatial Light Modulator

speed. In its simplest form, it consists of a photocathode, microchannel plate (MCP), a planar acceleration grid and an electro-optic crystal plate (see Figure 1.1). The crystal has a high resistivity dielectric mirror on the side that faces the grid, and a conducting electrode on the other. The voltages V_{MCP} , V_{grid} and V_b are external to the device.

In the electron-deposition mode, the write beam (coherent or incoherent light) incident on the photocathode creates an electron image which is amplified by the MCP and proximity focused onto the dielectric mirror. The resulting spatially varying electric field modulates the refractive index of the crystal. Thus, the

readout light which makes a double-pass through the crystal is phase or amplitude modulated, depending on the crystal cut and readout scheme (polarization or interferometric) employed.

The image is erased by flooding the photocathode with light so that the electrons are removed from the mirror by secondary electron emission. Alternatively, the device can be operated in the reverse mode, in which the image is written by removing charge from the dielectric mirror surface by secondary electron emission and erased by adding charge to the mirror.

The high exposure sensitivity, high spatial resolution, long storage and cascadable nature of the MSLM make it an especially versatile spatial light modulator. Optically-addressed MSLMs have already been utilised in adaptive phase compensation [5], optical image processing (such as contrast reversal, addition, subtraction, thresholding, edge enhancement and logic operations) [4]. More recently, an electron-beam addressed (E-beam) MSLM has been developed and used in matrix multiplication, with current research focused on implementing an optical inference machine for symbolic processing [6]. In the E-beam MSLM, the photocathode and the input light source are replaced by an electron gun whose beam is focused directly onto the MCP input face; images are written by controlling the electron beam using deflection grids.

A summary of the standard MSLM operation is included in Chapter 2, but no attempt is made to present MSLM theory in complete detail. For more complete coverage, the reader is referred to References [1]-[7].

The main focus of this thesis is to describe and analyse the techniques that can be used for converting the periodic write/readout intensity characteristic of the spatial light modulator to the desired switching characteristic. These techniques

include:

- converting the standard device to a Fabry-Perot MSLM
- operation in a closed loop feedback configuration
- operation in the electron deposition/grid stabilized mode
- operation in the real-time nonlinear hard-clipped thresholding mode

The standard MSLM and the Fabry-Perot MSLM are presented in Chapters 2 and 3 respectively. Both devices are characterised; device parameters, operating voltages, experimental set-ups and results are given. In addition, Chapter 3 includes a discussion of the criteria for image processing with a Fabry-Perot MSLM. Chapter 4 analyses the operation of an MSLM in the optical feedback configuration for achieving crystal compensation and bistable operation. In Chapter 5, the switching characteristic for the MSLM is defined and the intrinsic and variable gamma modes of operation characterised. Chapter 6 describes possible optical information processing applications in general, and in particular, the experimental results of real-time halftone screen processing using a standard MSLM. Finally, Chapter 7 summarises the important results of this thesis work and includes recommendations for future work.

Chapter 2

Standard MSLM

Substantial research has been devoted to the development and improvement of the standard MSLM. The theory of MSLM operation is well understood and is presented elsewhere in great detail. The following section contains a qualitative summary of the device description and theory necessary for the purposes of this thesis.

2.1 Device Description

As mentioned in the introduction, the standard microchannel spatial light modulator consists of a photocathode, microchannel plate (MCP), acceleration grid and electro-optic crystal plate.

The choice of input window material and photocathode determine the write light wavelength response of the MSLM. At present, there exist multi-alkali photocathodes for wavelengths spanning 110nm to 1200nm. As discussed later, the MCP input face can be excited directly using UV light, but there exists no photocathode that can be used in the infrared region above 1200nm.

A theory of the microchannel plate can be found in an article by Wiza [8]. The MCP (thickness about 0.5mm) is an array of semiconducting-glass pores (diameter about 10 μm) oriented parallel to each other. Each channel can be considered to be a continuous dynode electrode multiplier with a gain factor of $10^3 - 10^4$. Parallel electrical contact to each channel is provided by deposition of a metallic coating, usually nichrome or inconel, on the front and rear surfaces of the MCP, which then serve as input and output electrodes respectively. The MCP is spatial resolution limited only by the channel dimensions and spacings. Originally developed as an amplification element for image intensification devices, MCPs have direct sensitivity to charged particles (electrons and positive ions) and energetic photons (*UV*, *EUV* and *X-ray* wavelengths). During operation, the important parameters are the MCP bias voltage (V_{MCP}) and the strip current.¹ The gain of the MCP is characterised by the bias voltage; while the response time of the MCP² is inversely proportional to the strip current.

While the presence of an acceleration grid is not strictly necessary, it greatly enhances the MSLM operation. The presence of an acceleration grid biased positively (V_{grid}) with respect to the MCP output electrode (V_{MCPout}) serves two purposes. Firstly, the grid accelerates the electrons emitted from the MCP towards the crystal, so that a larger fraction of the ejected electrons contribute to the primary current density and good proximity focusing is maintained between the grid and crystal. Secondly, the grid increases the energy of the electrons ejected by the MCP by an amount eV_{grid} . Thus, the electron removal process is much more efficient in a device with a grid. Attractive features for an acceleration

¹current through the microchannel plate

²the response time is proportional to the channel recovery time (time to replenish charge after firing)

grid are high electron transmittance (large open area), high tensile strength, low elasticity and low secondary electron emission yield. A metal mesh (copper or nickel) spotwelded to a support ring is usually used.

Dielectric mirrors consist of alternating layers of high and low refractive index dielectric materials. Ideally, the dielectric mirror should have high damage resistance under electron beam bombardment, high reflectivity at the readout wavelength and low vapor pressure.

The electro-optic crystal plate is the most important component determining the MSLM performance. The mechanism for modulating the readout light making a double-pass through the crystal is the linear electro-optic effect induced by an electric field between the spatially-varying charge density on the dielectric mirror and the transparent conducting electrode. For Pockels effect crystal, the induced phase change $\phi_{x'}$, $\phi_{y'}$ in the crystal is proportional to the surface charge density, $\sigma(E)$ (where x' and y' are the induced crystal axes). The exact expression depends on the type and cut of the crystal employed. For example, in oblique-cut LiNbO₃, the electrically-induced phase retardation Γ is of the form,

$$\Gamma = \phi_{y'} - \phi_{x'} = \frac{\pi\sigma(E)}{\lambda C} (n_o^3 r_{y'} - n_e^3 r_{x'}) \quad (2.1)$$

where λ is the wavelength of the exposing light, C is the average capacitance per unit area of the crystal, n_o and n_e are the ordinary and extraordinary refractive indices respectively, and $r_{x'}$ and $r_{y'}$ are effective electro-optic coefficients. The amount of charge density $\sigma(E)$, or equivalently the voltage across the crystal, V_s , uniquely determines the amount of phase modulation. When such a crystal is read out between crossed polarizers, the transmittance of the crystal is given by,

$$T = \frac{I_r}{I_i} = \sin^2 \left(\frac{\Gamma(E)}{2} \right) \quad (2.2)$$

where the phase retardation Γ is a function of input exposure E . The crystal voltage that is needed to produce a phase modulation of π radians with a double-pass through the crystal is referred to as the reflection halfwave voltage, $V_{\pi R}$, corresponding to the reflection halfwave surface charge density, $\sigma_{\pi R}$. Devices to date have used z -cut LiNbO₃ for phase modulation and 55°-cut LiNbO₃ for intensity modulation. This 55° crystal cut optimises the phase retardation yielding a minimum calculated $V_{\pi R}$ of 1250V. An alternative way of expressing the modulator transmission in terms of the crystal voltage is,

$$T = \frac{I_r}{I_i} = \sin^2 \left(\frac{\pi V_z}{2V_{\pi R}} \right) \quad (2.3)$$

2.2 Principles of Operation

The voltage across the crystal, V_z , and the voltage across the vacuum gap, V_g , as shown in Figure 2.1 are related to the externally controlled voltages, V_b and V_{grid} , by:

$$V_b = V_g + V_z + V_{grid} \quad (2.4)$$

The effects of the ballast resistors, R_{grid} and R_b , are neglected. The gap is defined as the space between the grid and the dielectric mirror.

Changes in the externally applied voltages, V_b or V_{grid} , are divided capacitively between the gap and the crystal. However, since the crystal capacitance is much greater than the gap capacitance, the changes in voltage are coupled to a large extent across the gap. The crystal voltage, V_z , is therefore determined only by the charge density already present on the surface of the crystal.

Primary electrons are ejected from the MCP, accelerated by the grid, strike the dielectric mirror surface and generate secondary electrons. For sufficiently

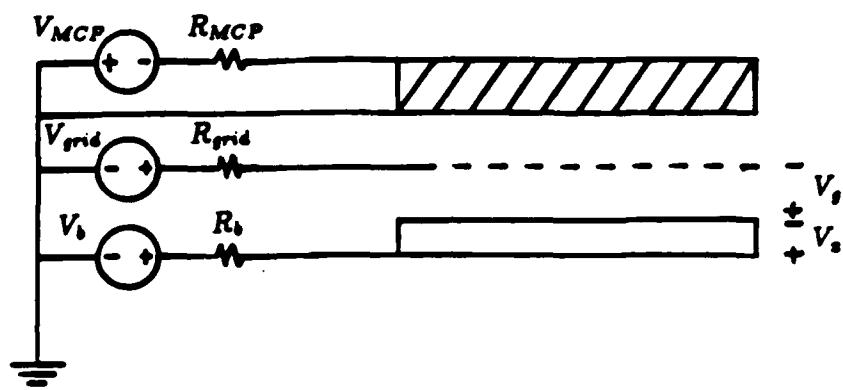


Figure 2.1: MSLM internal configuration

high primary electron energy, each primary electron generates more than one secondary. These secondary electrons are then collected by the mirror, or the grid, depending strongly on V_s .

For high values of V_s , all the secondary electrons are attracted back to the mirror, and charge accumulates on the mirror surface at some maximum rate. If neither V_b nor V_{grid} are adjusted, the accumulation of electrons will cause V_s to decrease. As V_s falls below about 50V, some secondary electrons will have sufficient energy to traverse the gap and be collected by the grid, and the rate of accumulation of charge on the mirror will decrease. This accumulation will continue until the number of primary electrons striking the mirror is equal to the number of secondary electrons collected by the grid, and V_s has reached some equilibrium value. This equilibrium voltage is usually a few volts positive, but for all practical purposes, is assumed to be zero volts; therefore, the grid and the surface of the mirror are assumed to be at the same potential in equilibrium.

If V_s is initially slightly below the equilibrium gap voltage, the generated secondary electrons will be collected by the grid, resulting in a net loss of electrons from the mirror until V_s is once again in equilibrium.

When there exists a large negative field across the gap, the energy of the primary electrons will be greatly reduced. This results in a much greater number of primary electrons striking the mirror than secondary electrons generated and collected by the grid. Thus, there will be an accumulation of electrons on the mirror which will cause V_s to decrease until the repulsive field is so large that the primary electrons can no longer traverse the gap. The system is now in *LOCKOUT*: the value of V_s at which this occurs is known as the *lockout voltage* and is the voltage below which no primaries have sufficient energy to reach the

mirror and are all collected by the grid. The crystal surface charge density can now only be changed by manipulating the external voltages, V_{grid} or V_b .

2.3 Linear Operating Modes

In a linear mode of operation, the incremental surface charge density $\sigma(E)$ deposited on the crystal is proportional to the exposure E such that,

$$\sigma(E) = \left(\frac{\eta e G}{h\nu} \right) E \quad (2.5)$$

where η is the photocathode quantum efficiency, e is the electronic charge, G is the MCP gain, h is Planck's constant and ν is the frequency of the write light. Therefore, for a Pockels effect crystal, the induced phase retardation in the crystal will also be proportional to E .

From the discussion in the previous section, it is now apparent that there are two such modes of operation: 1) write by electron deposition and 2) write by secondary electron emission (electron depletion).

In the electron-deposition mode, the gap voltage must be maintained at a high enough voltage level to ensure proximity focusing. Images may then be written by exposing the photocathode to some input I_w and integrating the charge on the mirror for a sufficient length of time to obtain a spatially-varying Γ proportional to I_w (eqns. 2.1 and 2.5). This is accomplished by first biasing V_{grid} up to 2kV. Then, the background level is adjusted by setting V_b to some bias level V_{bias} (dependent on the particular crystal), and flooding the photocathode with light until equilibrium is reached and there is a uniform charge distribution on the mirror. Next, V_b is set to some write voltage, V_{bw} and the photocathode exposed to the input image for an appropriate write time. The write process is terminated by turning off the

write light. The crystal may be erased by flooding the photocathode with light and depositing a uniform charge distribution as before, then lowering V_b to the appropriate level with the light still on.

The electron-depletion mode requires more sophisticated control of V_b . The grid is once again biased up to 2kV and the background level is adjusted as before. Then, with the write light on, V_b is ramped downwards at a preset rate $\dot{V}_b = dV_b/dt$ such that no areas of the mirror will be locked out. Electrons are therefore removed from the crystal surface in amounts proportional to I_w . Once again the write process is stopped by turning off the write light and stopping the ramp. This image is erased by raising the crystal voltage to some V_{bias} and depositing a uniform amount of electrons on the crystal surface by flooding the photocathode with light.

2.4 Experimental System

2.4.1 The Hamamatsu MSLM

The standard MSLM used for this thesis work was a vacuum-sealed MSLM fabricated by Hamamatsu Photonics K. K. (Model X-1699, MM171) [9]. The device structure is shown in Figure 2.2. The dimensions of this MSLM are 52mm in diameter and 80mm in length. The device employs a multi-alkali photocathode (type S-20) and an approximately 50 μm thick, 55°-cut LiNbO₃ crystal. Electron optics, in the form of the accelerating and focusing grids G1 and G2, are employed for transferring the electron image onto the MCP. The output window is coated with an anti-reflection coating on the outer side, whilst the inner side has a transparent coating which is both anti-reflecting and conducting (serves as the

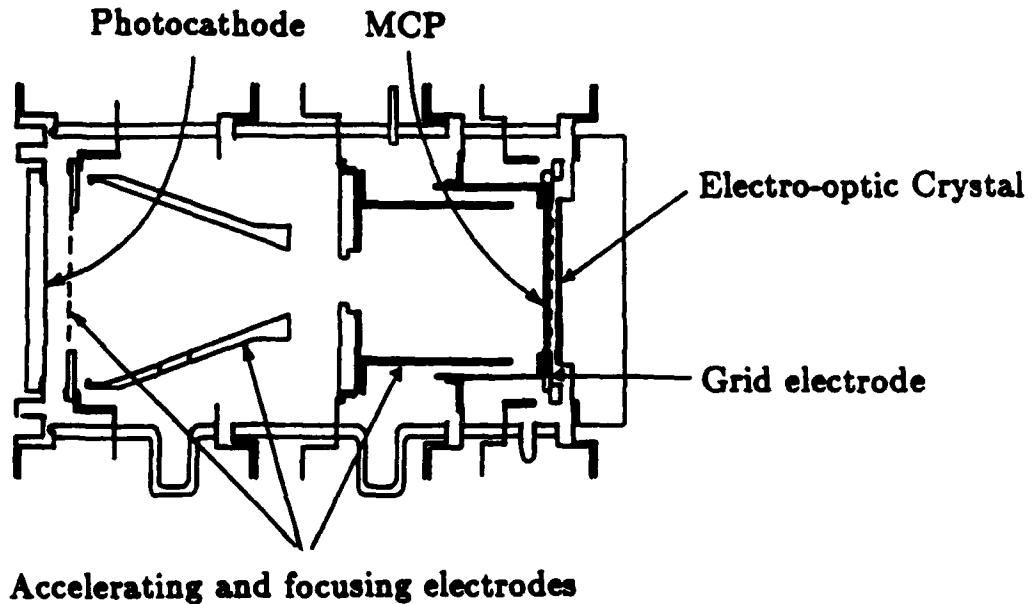


Figure 2.2: Vacuum-sealed standard Hamamatsu MSLM

crystal electrode). The crystal has a dielectric mirror on the MCP side and an anti-reflection coating on the other side, and is bonded with epoxy to the output window. A material with high secondary electron emission is evaporated on top of the dielectric mirror to improve the MSLM response time. The effective diameter of the crystal is about 18mm. The baking, cleaning and assembly procedure of this device is unknown. The vapor pressure is assumed to be well below 1×10^{-7} torr when sealed, to enable operation of the photocathode. Further information about the components and the device architecture is unavailable.

2.4.2 Power Requirements

Typical operating voltage requirements for the Hamamatsu MSLM and the

Electrode	Symbol	Voltage (kV)	Supply
Photocathode	V_{PC}	-3.0	Bertan Model 205A-05R
Focusing and Acceleration grids	V_{G1}	-1.5	MSLM Controller # 2, V_G
	V_{G2}	-2.7	MSLM Controller # 1, V_M
Anode (MCP _{in})	V_{anode}	-1.0	Bertan Model 205A-05R
MCP _{out}	V_{MCPout}	0	Grounded
Grid	V_{grid}	+1.0	Bertan Model 205A-05R
Crystal	V_b	0 - 3.0	MSLM Controller # 1, V_G, V_b

Table 2.1: Power requirements for Hamamatsu MSLM (Strip current 10 μ A)

high voltage power supplies actually used are shown in Table 2.1. Three Bertan high voltage power supplies (Model 205A-05R) were used to bias the photocathode, MCP and grid. The remaining power requirements were supplied by two MSLM Controllers (Model 1201/2), labelled # 1 and # 2, that were custom-made by the Massachusetts Manufacturing Corporation. Each controller has three floating *dc* high voltage, low current supplies. The controller also has a grounded crystal voltage supply which was designed specifically for biasing the crystal electrode of an MSLM. The crystal supply has both *dc* and framing modes. In the *dc* mode, the supply output is controlled by one of two front-panel controls V_b and V_{th} , which are used to set write and threshold voltages, respectively. This mode allows manual operation of the MSLM. When operated in the framing mode, the supply output follows the waveform of Figure 2.3, either single triggered or continuously cycled. The voltage levels and timing of indicated portions of the waveform can be controlled independently, and parts or all of the waveform can be generated. The controller also provides low voltage, AB (15V) logical outputs which were used to synchronise write and erase lights with different parts of the waveform. The MSLM controller was able to provide all of the necessary voltage waveforms required for the work herein.

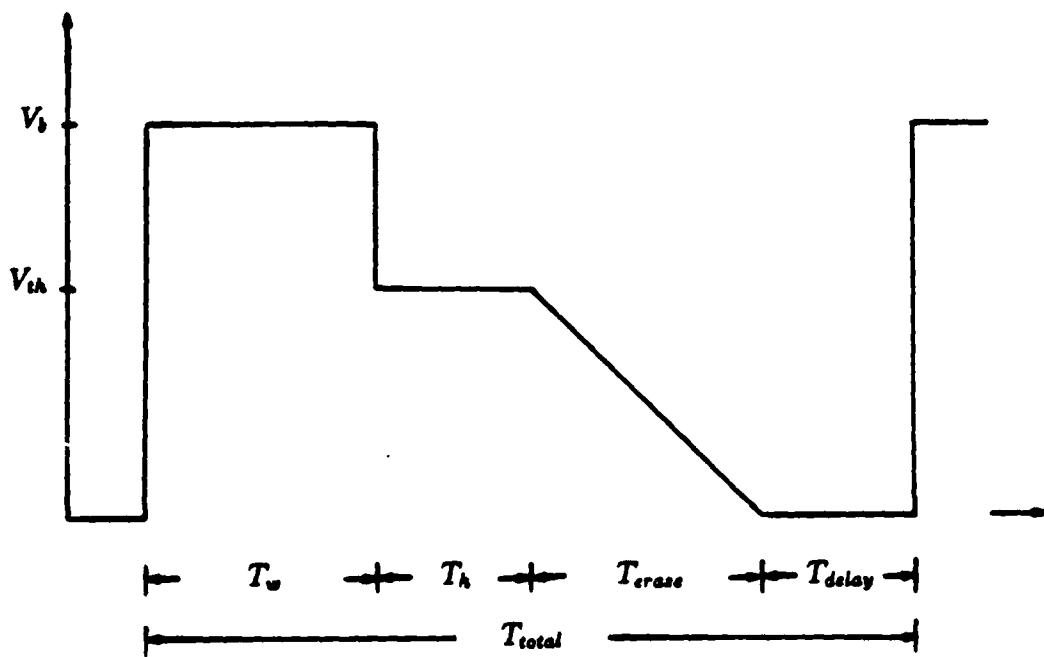


Figure 2.3: MSLM controller timing waveform

In order to protect the power supplies in the event of accidental arcing between MSLM components, ballast resistors were placed in series with the power supplies. Drainage resistors were placed in parallel with the Bertan supplies to accelerate the decay of their large internal capacitances. A capacitor was also added to the crystal electrode to eliminate rapid high voltage transients which occur during supply turn-on and turn-off. The MCP input and output currents were monitored during operation. Figure 2.4 shows the electronic biasing box that was built; great care was taken to ensure that all high voltage connections were spaced far apart and the aluminum box grounded.

2.4.3 Optical System

The optical system used to write and to readout information from the Hamamatsu MSLM is shown in Figure 2.5 . The crystal is read out between crossed polarizers oriented 45° with respect to the crystal axis. Spatial filtering in the transform plane removed any undesired reflections. The readout optics had a magnification of about 2.25. The output image was monitored using a Panasonic TV camera (Model WV 1550) and TV monitor; this allowed simultaneous controlling of the MSLM power and monitoring of the output image. The write light source was a 25 Watt Soft Light incandescent light bulb with a red filter placed in front to disperse the light. The write light was collimated as shown in the write optics and used to illuminate an image transparency in the input object plane. The input transparency could be translated in the $x - y$ transverse plane with about 0.5mm accuracy and along the z -direction (optic axis) with about 0.1mm accuracy. This allowed accurate positioning and focusing of the input image on the face of the photocathode; the write imaging optics had unity magnification. A red LED

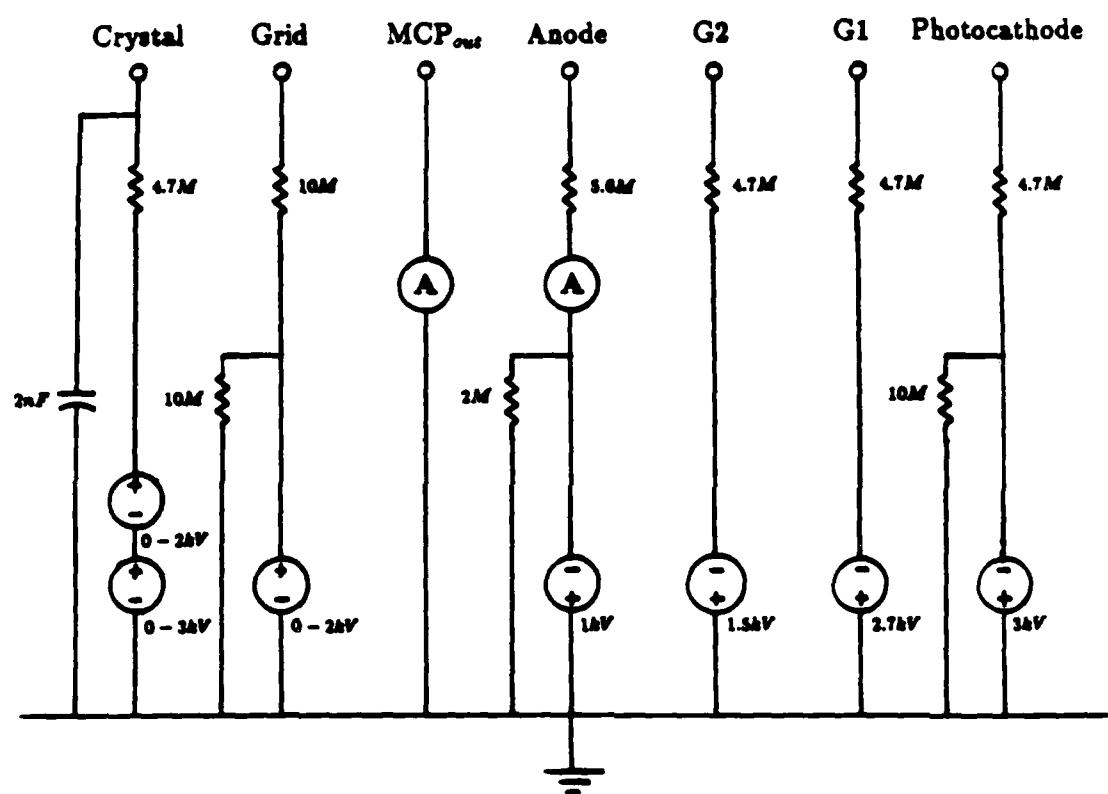


Figure 2.4: Standard MSLM biasing box

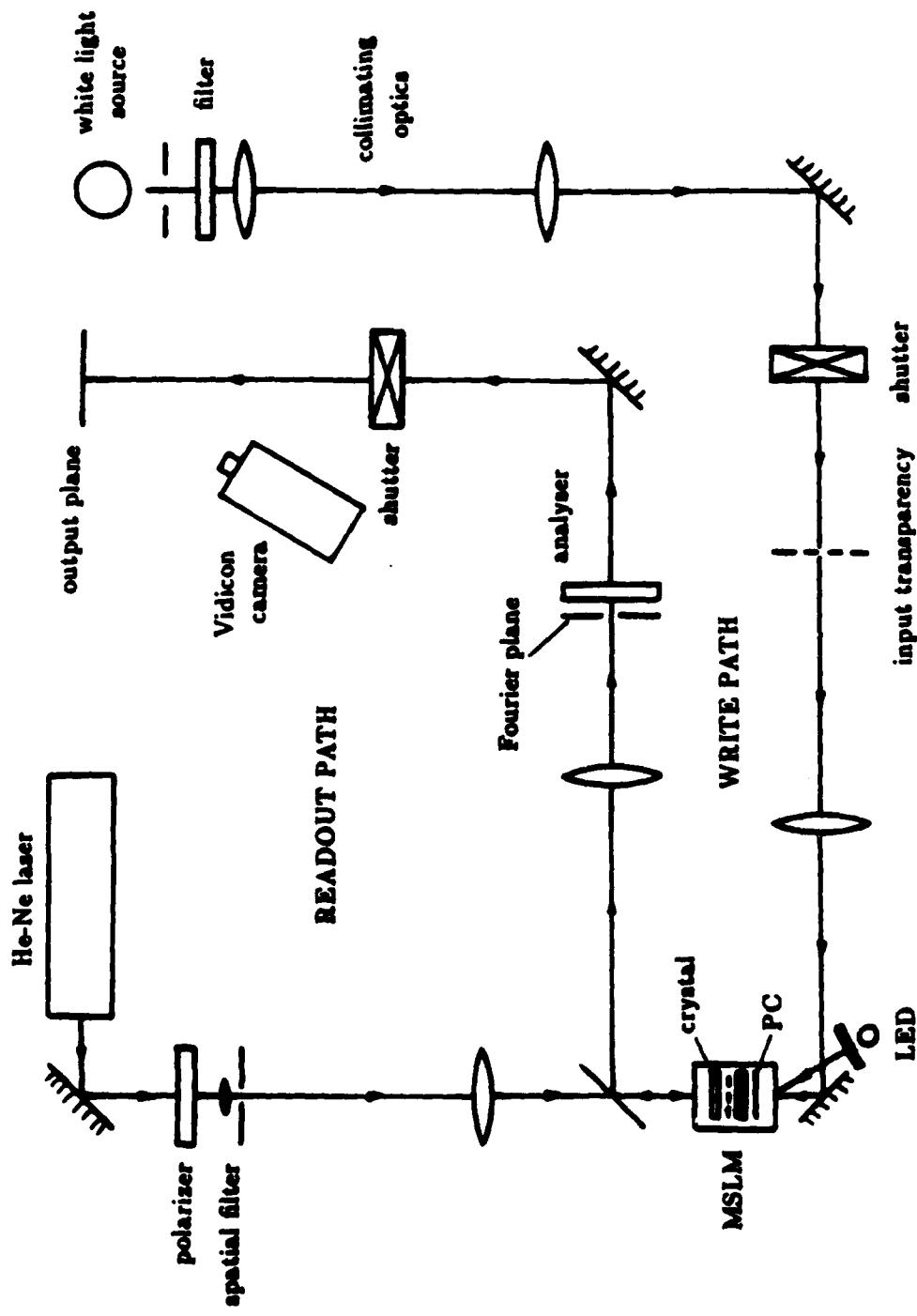


Figure 2.5: Optical configuration for Hamamatsu MSLM

placed behind a dispersive red filter, was employed to flood the photocathode with light during erasure and biasing. The write input image was shutter controlled while the erase LED was powered on and off as needed. The input transparencies were fabricated using 35mm Polaroid Polapan B & W continuous-tone film and mounted in slide holders. The image of a 1951 Air Force resolution chart was also used as an input pattern. The Air Force resolution chart itself was made by Opto-Line Corporation and consisted of an etched chrome layer on a quartz substrate (the transmittance of the opaque layers was less than 0.1%).

The visible photocathode was found to be extremely sensitive to any stray light in the room. While this sensitivity is desirable in the MSLM, input images written onto the device were being *smeared out* by very low background light levels. To resolve this problem, the entire write path was totally enclosed and the photocathode completely isolated from any stray room light, readout laser scattering and even write light reflections within the write path. This was accomplished quite simply using cardboard boxes, baffles, tubing and substantial amounts of duct tape.

2.4.4 Output Devices

Output images were photographed using Polaroid type 55PN film with the exposure time controlled by a shutter placed in front of the laser. For those cases where the intensity in the output plane was measured, a Hamamatsu Si PIN diode biased in its linear region was used. The voltage level V_{detout} , proportional to the output image intensity, I_o , is equal to the measured voltage level, $V_{detmeas}$, minus the voltage due to the detector dark level and the ambient light level, $V_{detdark}$,

$$V_{detout} = V_{detmeas} - V_{detdark} \quad (2.6)$$

Thus, before any data is taken in a given run, a reading of $V_{detdark}$ is taken. All of these voltage readings are taken after a sufficient settling time and are assumed to be time-averaged.

2.5 Device Performance

Since the Hamamatsu MSLM was vacuum-sealed and the internal components of the device unchangeable, great care was taken to avoid any destructive operating regions. The maximum possible grid voltage, V_{grid} , was found to be 2.3kV ; at this voltage, the force between the grid and the MCP output face was large enough to stretch the grid and touch the MCP, resulting in arcing and damage of the MSLM power supply being used for the grid. The grid was thereafter set at $+1.0\text{kV}$ which allowed sufficient proximity focusing and convenient manipulation of V_t during writing. The crystal voltage V_t was generally limited to a range of $0 - 3.0\text{kV}$ during operation; voltages outside of this *safe* range resulted in damage to the crystal (see Appendix B). The MCP was biased at -1.0kV with a strip current of about $10\mu\text{A}$. The photocathode was biased at -3.0kV , and the optimal focusing grid voltages were $V_{G1} = -1.5\text{kV}$ and $V_{G2} = -2.7\text{kV}$. V_t was controlled by a floating supply to set the bias level and the MSLM timing waveform to produce the required high voltage profiles for operation.

A number of idiosyncrasies were observed while operating this device and are believed to be related to some device modifications made on the original MSLM prototypes. These are discussed at length in Appendix B. These peculiarities were superimposed on the output images and can be observed in the photographs taken in the output plane. While these effects are aesthetically unattractive, they are not universal to MSLMs and are neglected for the purposes of this thesis.

Recall that the modulator has a sine-squared dependence with respect to the voltage across the crystal V_z for an oblique-cut LiNbO₃ crystal read out between crossed polarizers, such that,

$$T = \frac{I_o}{I_i} = \sin^2 \left(\frac{\pi V_z}{2V_{zR}} \right) \quad (2.7)$$

where I_i is the input intensity, I_o is the reflected readout intensity and V_{zR} is the reflex halfwave voltage. A plot of T vs. V_z is shown in Figure 2.6. V_{zR} can therefore be measured quite simply by monitoring the intensity of the output image for a series of V_z settings when a constant write intensity is incident on the photocathode. After a sufficient settling time for each V_z setting, the gap voltage equilibrates to about 0V, and the voltage change in V_z couples entirely across the crystal.

When this measurement was attempted on the Hamamatsu MSLM, the voltage V_z was varied from -1.0 to +3.0kV with a V_{grid} of 1.0kV (that is, V_z from -2.0 to +2.0kV). A minimum in the output intensity occurred when $V_z \approx 0.9 - 1.0\text{kV}$, depending on how long the device had been operating. However, no maxima were observed and the output intensity still increased at either end of the allowed voltage range. The range of V_z was not increased any further for fear of detrimental effects to the crystal (Appendix B). The data obtained by this measurement was fitted to a sine-squared curve, and the mean square error minimised with respect to V_{zR} . This method yielded an estimate for V_{zR} of about 4.5kV for the LiNbO₃ crystal in the Hamamatsu device. This does not agree with the predicted value of 1250V for a 55°-cut LiNbO₃ crystal. An explanation for this discrepancy requires more information about knowledge about the crystal and the device construction. Further investigation and actual measurement of V_{zR} would have resulted in destructive operation of the device and was therefore not

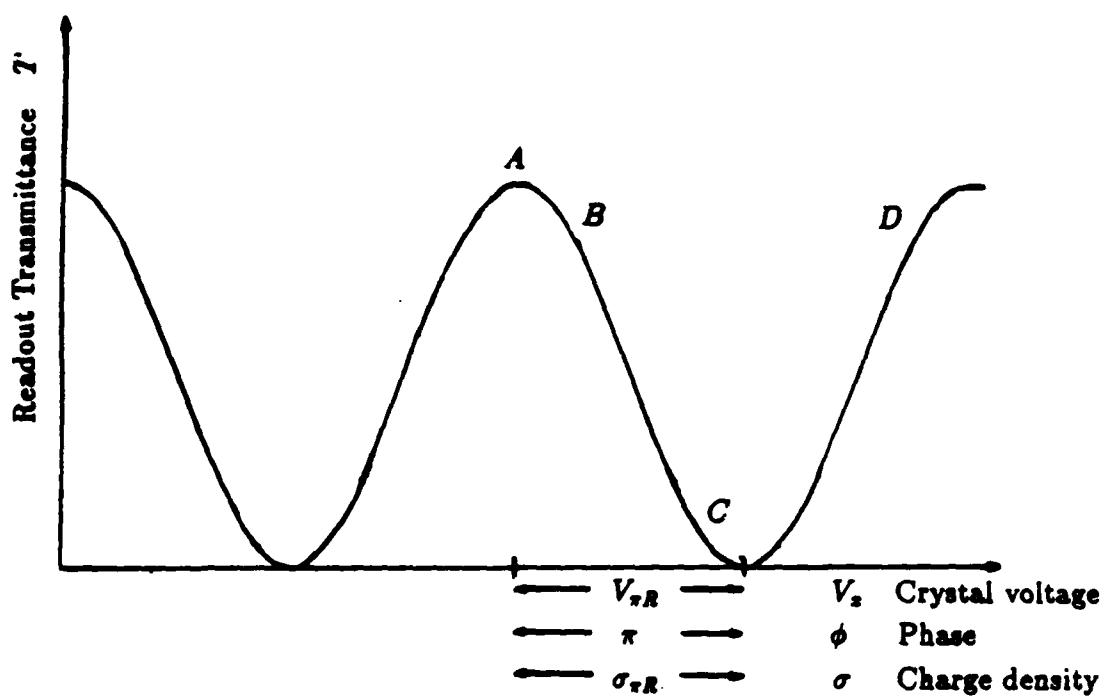


Figure 2.6: Readout Characteristic for a standard MSLM with Pockels effect crystal

attempted.

This standard MSLM was operated in what is believed to be the σ_0 to $\sigma_0 + \sigma_{\pi/2}$ region of the characteristic modulator readout curve (corresponding to V_z from -1.0 to +2.0 kV).

Two important parameters for assessing the performance of an MSLM are the framing rate and the spatial resolution. The cycle time of a device is defined as the sum of the time required to deposit σ_π charge, t_{ds} , and the time required to remove σ_π charge, t_{rr} ,

$$t_\pi = t_{ds} + t_{rr} \quad (2.8)$$

The framing rate $1/t_\pi$ is dependent on the input intensity, photocathode efficiency, MCP gain, MCP strip current (high strip current MCP's have fast recovery times), the energy of the primary electrons through V_s , the halfave charge density σ_π , the spacing between components and device structure. No attempt was made to measure the maximum framing rate of this device since it would require V_z of $V_0 + V_{zR}$ (> 4.5 kV), a much higher voltage on the grid and saturation of the photocathode (this tends to degrade the photocathode sensitivity). Typical erase times ranged from 1 - 60 seconds for erasure from D to C (Figure 2.6).

The spatial resolution characteristics of a device, or optical component are usually evaluated in terms of the Modulation Transfer Function (MTF). When the input signal is an image, as in the MSLM case, the abbreviation MTF_u is used to indicate a light-to-light coherent modulation transfer function. The normalized MTF_u is defined as,

$$MTF_u = \frac{1}{m} \frac{I_{max}(\nu) - I_{min}(\nu)}{I_{max}(\nu) + I_{min}(\nu)} \quad (2.9)$$

where $I_{max}(\nu)$ and $I_{min}(\nu)$ are the maximum and minimum intensity values at spatial resolution ν in the coherent readout beam and m is the spatial modulation

index of the input write light. The MTF_u was experimentally determined by using the image of a 1951 Air Force resolution chart, and projecting it onto the face of the photocathode with the write optics. Great care was taken to focus the resolution chart using a precision translator along the optic axis, and to prevent any stray light from reaching the photocathode. The modulation index of the input image, m , was assumed to be about unity for all the input frequencies that were used.

The entire crystal was first biased at C (this corresponds to the minimum modulator transmittance) and the image of the resolution chart written by electron deposition with an average charge density of $\sigma_0 + \sigma_{\tau/4}$. This was accomplished by writing for a time t_w , which is the time required to deposit enough charge on the crystal to reach D with a uniform write light intensity (zero spatial frequency); this must be the same write light that is used to illuminate the resolution chart. Figure 2.7a is a photograph of the modulated output image written as above. The negative from this photograph was mounted on a glass plate and scanned using a Jarrel-Ash # 21-050 Microphotometer. After the background transmittance of the negative/glass plate and the recording characteristics of the film were taken into account, the values of $I_{max}(\nu)$ and $I_{min}(\nu)$ obtained from the microphotometer were used to calculate MTF_u using Equation 2.9. Care must be taken to expose the film such that all information is recorded in its linear region. This standard MSLM had a measured spatial resolution of about 4 lp/mm at 50% modulation according to the above procedure.

Figure 2.7b is a photograph of the resolution chart written by electron removal. In this case, the crystal is first biased at D , then with the resolution chart incident on the photocathode, the voltage V_b is ramped downward to C at a rate \dot{V}_b ; this rate corresponds to the rate required to have a charge density σ_0 on the crystal at

CHAPTER 2. STANDARD MSLM

35

the end of the ramp when a uniform write light (zero spatial frequency) is used.



(a)



(b)

Figure 2.7: Resolution chart image by: a) electron deposition, and b) electron depletion

Chapter 3

Fabry-Perot MSLM

3.1 Device Description

The Fabry-Perot version of the MSLM (F-P MSLM)[10] employs a crystal that functions as an electro-optically tunable, Fabry-Perot etalon when electrons are deposited or removed from its surface. To fabricate such a crystal, standard dielectric mirrors are deposited on both surfaces of the crystal and then a transparent conducting overcoat is deposited on the readout surface of the crystal.

For the *z*-cut LiNbO₃ device operated in the reflex mode, the phase retardation Γ which results from the electrically induced refractive index change Δn , is, to a good approximation, proportional to the surface charge density $\sigma(E)$ and to the modulator write light exposure E . Γ is given by:

$$\Gamma = \phi_z = \phi_y = \frac{2\pi n_0^3 r_{13} \sigma(E)}{\lambda C} \quad (3.1)$$

where ϕ_z and ϕ_y are the phase changes along the respective crystal axes, λ is the wavelength of the exposure light and V_z is the voltage across the crystal.

For a Fabry-Perot etalon with surfaces of reflectivity R , it is well known that

the ratio T , of the total reflected intensity to the incident intensity, is given by,

$$T = \frac{I_r}{I_i} = \frac{4R \sin^2\left(\frac{\phi}{2}\right)}{(1-R)^2 + 4R \sin^2\left(\frac{\phi}{2}\right)} \quad (3.2)$$

This equation can be rewritten as,

$$T = \frac{I_r}{I_i} = \frac{F \sin^2\left(\frac{\phi}{2}\right)}{1 + F \sin^2\left(\frac{\phi}{2}\right)} \quad (3.3)$$

where the coefficient of finesse F is defined as,

$$F = \frac{4R}{(1-R)^2} \quad (3.4)$$

Figure 3.1 is a plot of T vs. ϕ ; V_s is given by Equations 3.1 and 3.2. Note that the reflected readout transmission of the F-P MSLM approaches zero when ϕ takes on integer multiples of 2π radians. The halfwave voltage V_r is once again defined as the voltage required for phase modulation of π radians. For z-cut LiNbO₃, the calculated V_r is,

$$V_r = \frac{\lambda}{2n_e^2 r_{13}} = 3100V \quad (3.5)$$

Notice that in Figure 3.1, $I_r \neq 0$ for $V_s = 0$; this would require that the crystal be polished exactly to the correct thickness and is unnecessary since I_r can be biased to any desired level by depositing charge on the crystal. An important measure of the sharpness of the Fabry-Perot fringes is the finesse \mathcal{F} , which is defined as the ratio of the separation of adjacent maxima to the full width at half maximum ($FWHM$) and is given by,

$$\mathcal{F} = \frac{\pi\sqrt{F}}{2} = \frac{\pi\sqrt{R}}{(1-R)} \quad (3.6)$$

Thus, a Fabry-Perot MSLM with high \mathcal{F} will have rapid drop-off in intensity on either side of the zero. The \mathcal{F} can be expressed in terms of ϕ , V_s or the induced

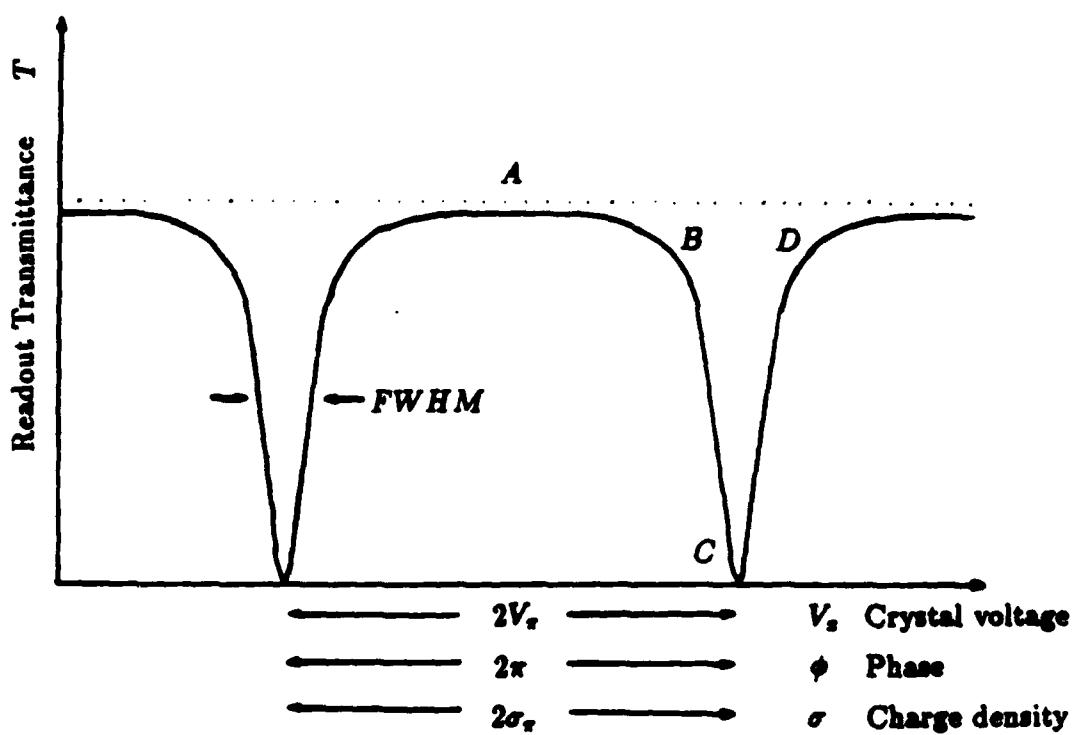


Figure 3.1: Readout characteristic for a Fabry-Perot MSLM

pathlength through the etalon, Δ ($\phi = \frac{2\pi}{\lambda} \Delta$),

$$\mathcal{F} = \frac{2\pi}{FWHM(\phi)} = \frac{2V_s}{FWHM(V_s)} = \frac{\lambda}{FWHM(\Delta)} \quad (3.7)$$

where λ is the wavelength of the readout light and $FWHM(\cdot)$ is the full width at half maximum expressed in terms of ϕ , V_s , or Δ (see Figure 3.1). Notice that the voltage swing required for a full light-to-dark transition (or vice versa) transition V_t is approximately equal to $FWHM(V_s)$ and thus,

$$V_t \approx FWHM(V_s) = \frac{2V_s}{\mathcal{F}} = \frac{4V_s}{\pi\sqrt{F}} \approx \frac{V_s}{\sqrt{F}} \quad (3.8)$$

Therefore the F-P MSLM only requires a voltage swing of $\frac{V_s}{\sqrt{F}}$ for full light-to-dark transition as compared to $V_{s,R}$ for a standard device. If the F-P is initially biased at its zero point, this represents a \sqrt{F} improvement over the speed of standard devices.

3.2 Linear Operating Modes

The high-finesse characteristics of the F-P MSLM lead to almost nonlinear operation when operated in its linear modes; this is particularly powerful for implementing nonlinear processing without any additional filtering. The electron accumulation and the electron deposition modes of the F-P device are similar to those discussed for a standard device. However, in its linear dynamic region, the F-P device is much more sensitive to the surface charge density on the mirror; thus, great control of the write light and external voltages is required for successful operation.

Typical electron deposition write mode is accomplished by biasing the crystal at A (Figure 3.1) by flooding the photocathode with light, setting V_t to point A

and allowing the device output to equilibrate ($V_s = V_{eq}$). The image is then written by setting V_s to point C and turning on the write light for sufficient exposure time t_w . This operation is more difficult with the F-P device since care must be taken to set V_s accurately during the write process to avoid over-accumulation of charge past C.

The electron-removal mode is somewhat simpler to implement experimentally and is accomplished by first biasing the crystal at D. Then, with the write light on, V_s is ramped downward from D to C volts at an appropriate rate \dot{V}_s . Since this mode only requires that the ramp \dot{V}_s and the write light be synchronised, the beginning and the end points of the operation are well defined.

3.3 Experimental System

3.3.1 Prototype Fabry-Perot MSLM

A vacuum-demountable F-P MSLM was designed, constructed and its performance evaluated. The details of the device design and assembly are included in Appendix A.

An MSLM requires a vacuum of at least 5×10^{-6} torr for the operation of the MCP and even lower for operation of the photocathode. The materials used to construct the MSLM must therefore have vapor pressure well below these pressures. All hardware components of the MSLM, in addition to the vacuum cell and system, were scrupulously cleaned during assembly or modification of the device. This was accomplished by rinsing in trichloroethane (for degreasing), acetone, methyl and a final wash in semiconductor grade acetone. Plastic gloves were used in all stages and dust was removed using nitrogen gas. Since the MCP

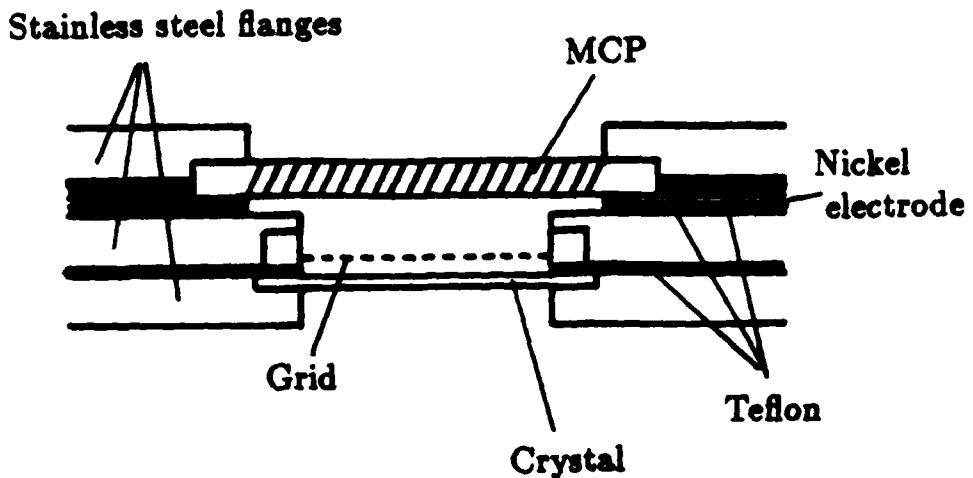


Figure 3.2: Prototype Fabry-Perot MSLM

was shipped from the manufacturer cleaned and under low vacuum, it was not subjected to cleaning. The crystal was also not cleaned, since this would damage the deposited coatings on its surfaces. Care was taken to ensure that all handling utensils were also cleaned. Good vacuum practice was found to be invaluable.

The F-P MSLM was mounted on a 152mm (6.0") flange which mated to a custom-made vacuum cell. The vapor pressure of the cell was maintained at about 2×10^{-6} torr by a 20 liter/s ion pump.

Figure 3.2 contains a diagram of the assembled F-P MSLM. Important features of the device components used are summarised in Table 3.1. Notice that the input window was made of S1 quartz and there was no photocathode; the MCP was therefore directly illuminated using a UV source. The high strip current MCP had a gain of about 10^3 and strip current of 280 μ A at -850V. The

Component	Description
Input Window	S1 quartz
MCP	<p><i>Galileo Electro-Optics Corp.</i> MCP # 7 · 204 - 010 Diameter: 25mm Channel diameter: 10μm Bias angle: 10° I_s (at 850V): 280 μA Gain (at 850V): 2×10^3 Vapor pressure: $\leq 5 \times 10^{-6}$ torr</p>
Grid	<p><i>ITT/Electro-Optical Prod. Div.</i> Nickel mesh 30 wires/mm Transmission: 60% Mounting ring i.d.: 18mm (0.7") Mounting ring o.d.: 22mm (0.876")</p>
Crystal	<p>z-cut LiNbO₃ Thickness 300 μm Diameter: 25mm (1.0") Flatness: $\pm \lambda/4$ Coatings by <i>Balzers Optical Group</i> Dielectric mirrors on both sides $\text{SiO}_2/\text{ZrO}_2$ (Top layer SiO_2) Reflection: 70% ± 4% Transparent conducting electrode ITO 200 Ω/ Transmission: $\geq 90\%$</p>
Output Window	S1 quartz

Table 3.1: Fabry-Perot MSLM components

grid consisted of 30 wires/mm, 60% transmission nickel mesh spot welded to a support ring with inner diameter 18mm (0.7"). The z-cut LiNbO₃ crystal plate was coated with dielectric mirrors on both sides and a conducting electrode on one side only. The dielectric coating consisted of alternating layers of SiO₂/ZrO₂ (top layer SiO₂) with a reflection R of $70\% \pm 4\%$ at a wavelength of 633nm. The conducting electrode coting was ITO (indium tin oxide), conductivity 200 Ω / and transmission $\geq 90\%$. The crystal plate was 300 μm thick and 25mm (1.0") in diameter. When read out with a coherent source, the crystal was found to have a single Haidinger interference fringe around its perimeter. This indicated crystal a plane-parallelism of about $\pm \lambda/4$. Teflon spacers were cut from 0.127mm (0.005") thick teflon sheet, which was found to be flexible enough to stay in position without warping and exerting pressure on the crystal or MCP.

The active diameter of the F-P MSLM with the configuration shown in Figure 3.2 is 18mm (corresponding to the active diameter of the grid). Electrical connections are made through high voltage feedthroughs in the 6.0" flange. The MCP-to-grid spacing was about 1.6mm, and the grid-to-crystal spacing was about 0.26mm for this design. The most significant contributor to the MCP-to-grid spacing was the thickness of the grid mounting ring.

3.3.2 Power Requirements

Typical operating voltage requirements for the F-P MSLM and the high voltage supplies used are shown in Table 3.2. The MSLM controller was once again used to control V_b , the crystal voltage. The protective ballast resistors and power supply connections are shown in Figure 3.3.

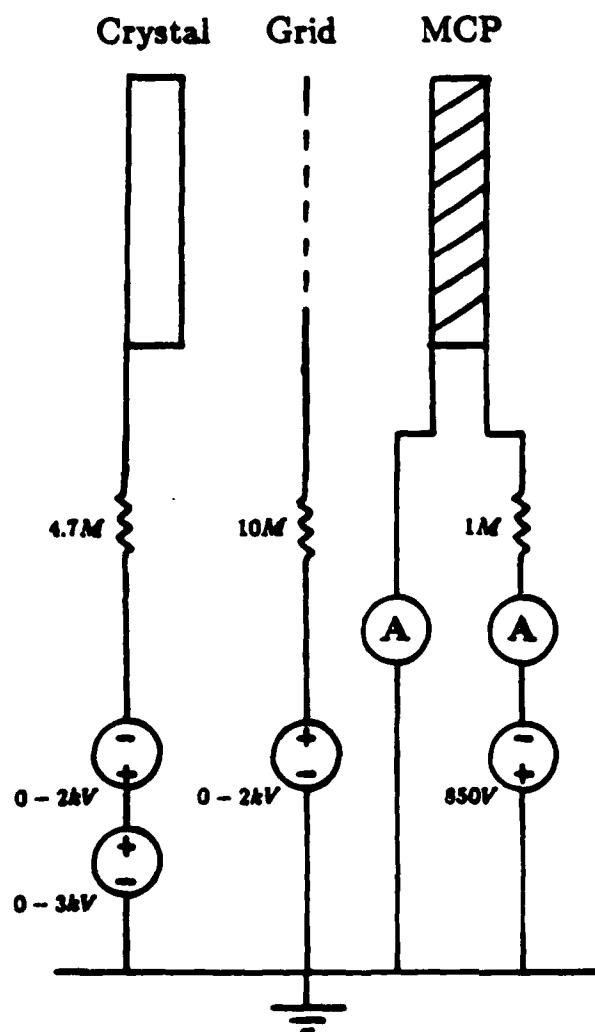


Figure 3.3: Fabry-Perot MSLM biasing box

<i>Electrode</i>	<i>Symbol</i>	<i>Voltage (kV)</i>	<i>Supply</i>
Anode (MCP _{in})	V_{anode}	-0.85	Bertan
MCP _{out}	V_{MCPout}	0	Grounded
Grid	V_{grid}	+1.20	MSLM Controller
Crystal	V_b	-1.0 - +2.0	MSLM Controller

Table 3.2: Fabry-Perot MSLM power requirements

3.3.3 Optical Systems

The optical configuration in Figure 3.4 was used to write and read out information from the F-P MSLM. The particular orientation of the MSLM was due to space constraints in positioning the vacuum cell. Notice that in this case the crystal was read out interferometrically so there were no polarizers in the the readout path. The spatial filter in the transform plane separated the reflections due to the output window surfaces from the modulated readout signal. The readout imaging optics had a magnification of 4. The collimating and imaging lenses were high quality Fourier transform quality lenses from Space Optics Research Labs. These lenses ensured excellent collimation of the readout light and accurate imaging of the output signal. The write optics consisted quite simply of a Phillips high pressure mercury spectral lamp as a UV source and a shutter to gate the light. The lamp was positioned as close as possible to the input window and isolated using a cardboard box to prevent stray UV from reaching the MCP when the shutter was closed. Output images were monitored once again with the video camera.

3.3.4 Output Devices

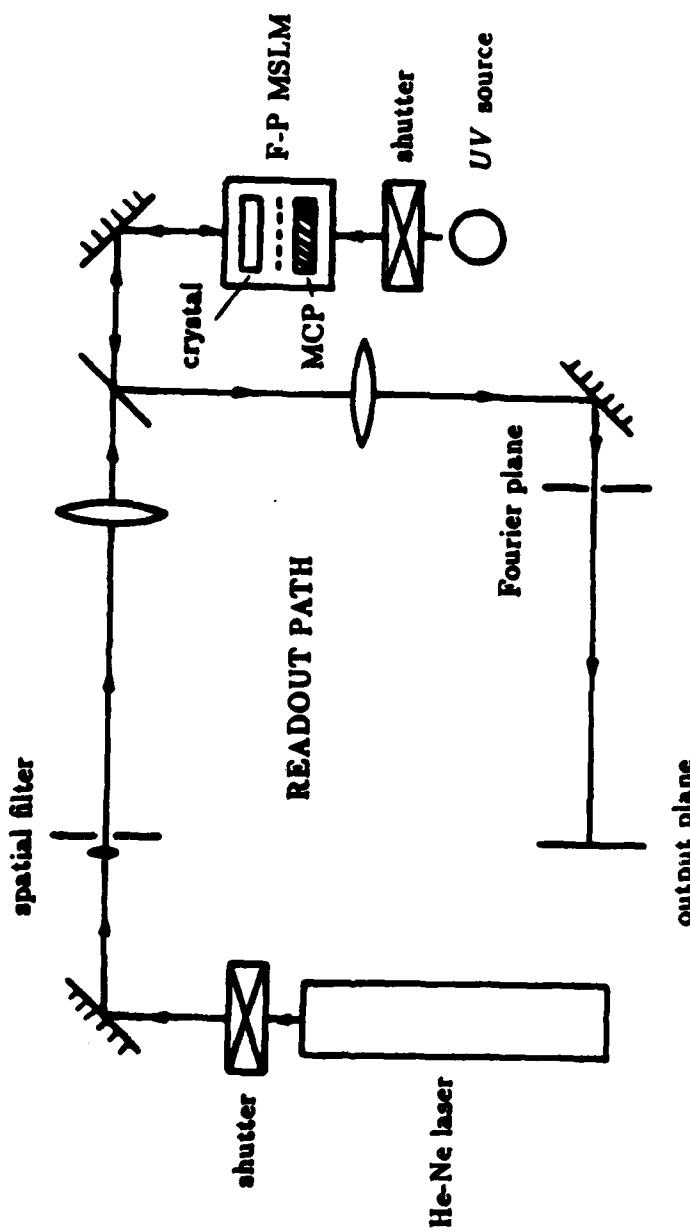


Figure 3.4: Optical configuration for Fabry-Perot MSLM

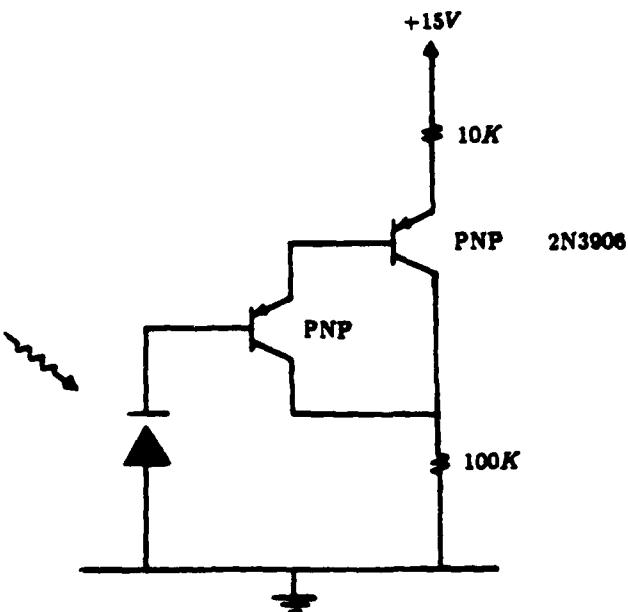


Figure 3.5: Detector amplifier circuit

Photographs of the readout signal were once again made with Polaroid type 55PN film exposed for a controlled time. Due to the nature of the F-P output, whenever the output intensity I_o was monitored using a detector in the output plane, the signal measured was localized by positioning a pinhole directly in front of the detector. As before, a Hamamatsu Si PIN diode was used together with the circuit shown in Figure 3.5 to amplify the detector signal. The voltage V_{detout} corresponding to I_o was once again,

$$V_{\text{detout}} = V_{\text{detmeas}} - V_{\text{detdark}} \quad (3.9)$$

where V_{detmeas} is the actual voltage measured by the detector and V_{detdark} is the voltage corresponding to the detector dark current and the ambient room light when the data was taken.

3.4 Device Performance

The reflectivity R of the dielectric mirrors deposited on the crystal surfaces and the plane-parallelism of the crystal surfaces proved to be the determining factors of the F-P MSLM performance. R determines the finesse \mathcal{F} of the etalon (Equation 3.6). By definition, the higher the \mathcal{F} , the narrower the $FWHM$ and the more sensitive the etalon becomes to deviations of the crystal surfaces from plane-parallelism.

The experimental F-P MSLM had a crystal coated with mirrors of $R = 0.7$. This corresponds to a finesse $\mathcal{F} \approx 8.76$. The crystal had a single Haidinger fringe around its 25mm perimeter before device assembly; this corresponds to crystal surface plane-parallelism of about $\approx \pm\lambda/4$.

The modulator was operated with an MCP bias of $-850V$ ($V_{anode} = -850V$) and strip current $I_s = 280\mu A$. The MCP was flooded with UV light and the grid voltage V_{grid} slowly raised to $+1.2kV$. This allowed the gap voltage V_g to equilibrate and reduced any risk to the grid and crystal. The crystal voltage V_b was varied between $-1.0kV$ and $2.0kV$ and the resulting modulation observed.

Figure 3.6 shows the output image for the crystal biased with $V_b = -0.95kV$. The dark area represents the part of the crystal where the etalon pathlength results in zero transmission (Equation 3.2). Unfortunately, this proved to be the widest area of the crystal to modulate concurrently. Lowering V_b resulted in inward movement of the modulation area towards the crystal center, whilst gradually becoming more circular in shape. Similarly, raising V_b resulted in the modulation area moving towards the edge of the crystal until the entire surface was finally biased bright.

No imaging could be accomplished using this experimental system, since the modulation area was not large enough. However, the high finesse operation of the

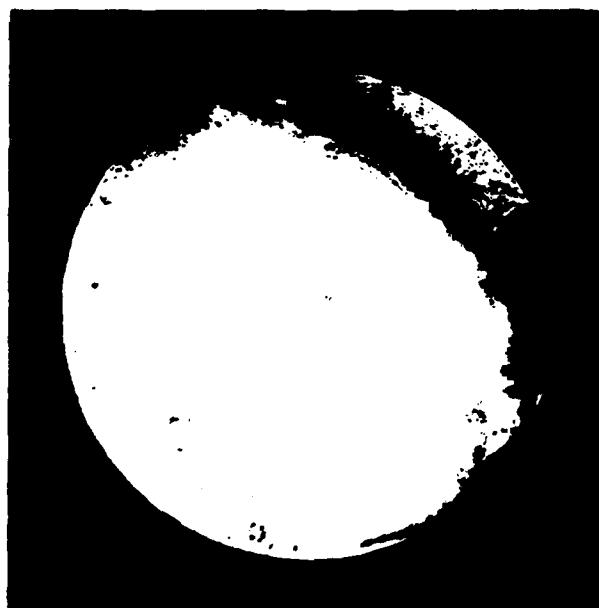


Figure 3.6: Fabry-Perot output image

F-P MSLM was characterised by positioning the detector/pinhole combination in the output plane in an area of the readout image that could be modulated, and recording the detector signal with a multimeter, oscilloscope or chart recorder.

Figure 3.7 is a plot of I_o (the intensity in the output plane), and the corresponding plot of the crystal voltage V_t . The Fabry-Perot MSLM readout characteristic was measured by ramping V_t downward from $+0.75\text{kV}$ to -1.0kV , while flooding the MCP with UV light. The removal of charge from the crystal results in modulation of the readout signal. The transition voltage V_t was measured to be about 700V . From Equation 3.8, V_s is estimated by $V_s = \sqrt{F}V_t = 3.9\text{kV}$. The maximum grid voltage V_{grid} was found to be about 2.1kV ; above which the electrostatic force between the grid and the MCP output face caused inelastic stretching in the center of the grid. Also, gap voltages above 1.0kV were avoided since these resulted in stress on the crystal and some warping of the crystal plate. The MCP was operated at -850V (as compared to -1.0kV , for the standard device) in an attempt to minimise the heating of the crystal due to the high strip current.

3.5 F-P MSLM Image Processing Criteria

The physical limitation on the finesse \mathcal{F} of a F-P MSLM that would be useful in image processing, is set by the deviation of the polished electro-optic crystal plate surfaces from plane-parallelism. The finesse \mathcal{F} was defined earlier as the ratio of the separation of adjacent maxima to the $FWHM$, and is expressed in terms of the mirror reflectivity R in Equation 3.6. Notice in Equation 3.7 that as \mathcal{F} increases, $FWHM$ decreases. By definition, the $FWHM(\Delta)$, expressed in terms of optical pathlength in the etalon, is a measure of the tolerance of the etalon to deviations from plane-parallelism of the crystal faces. That is, a region of the

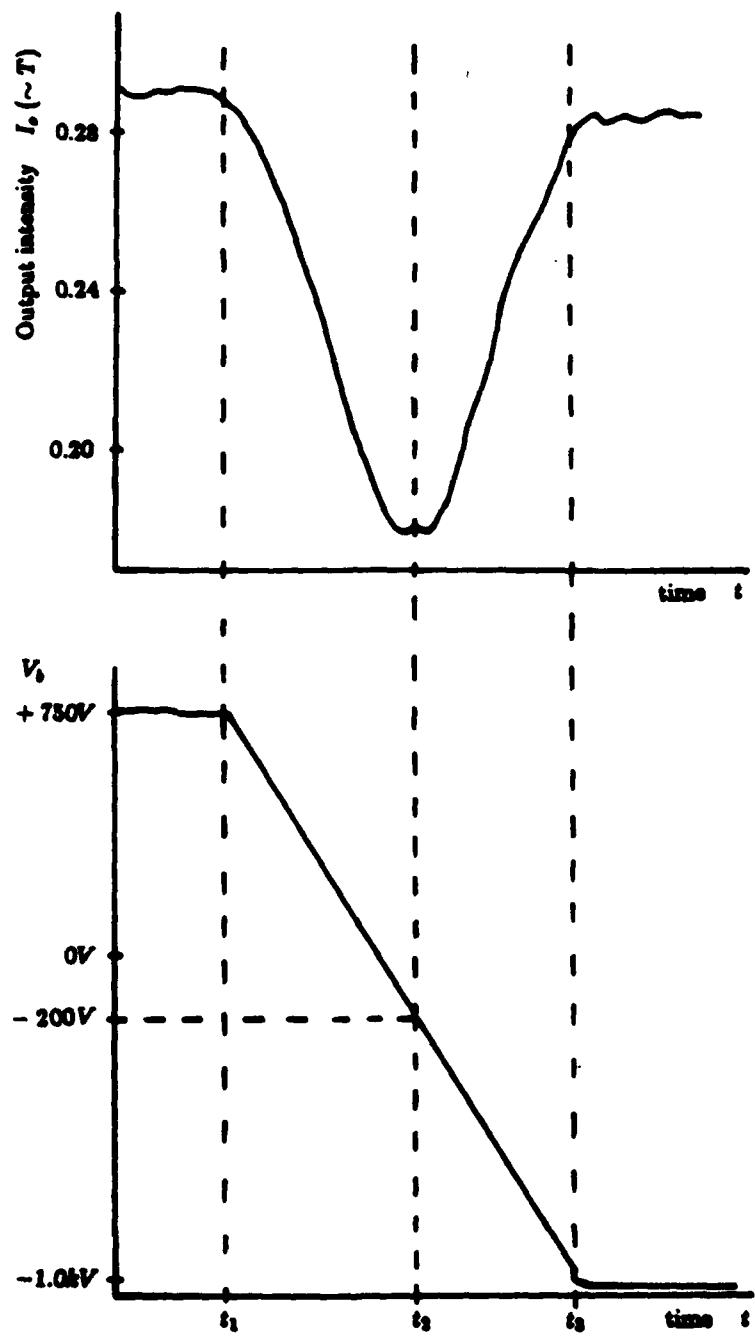


Figure 3.7: Experimentally measured Fabry-Perot MSLM readout characteristic

crystal with thickness l_s will modulate simultaneously with all other areas of the crystal whose thickness l satisfies the following,

$$l_s - \frac{FWHM(\Delta)}{2} \leq l \leq l_s + \frac{FWHM(\Delta)}{2} \quad (3.10)$$

This can be rewritten,

$$|l - l_s| \leq \frac{FWHM(\Delta)}{2} \quad (3.11)$$

If a measure of the crystal plane-parallelism a is defined as,

$$a \equiv |l - l_s| \quad (3.12)$$

where \bar{l}_s is the average crystal thickness, and l is the crystal thickness at any point of the crystal. For a significant area of the Fabry-Perot to modulate simultaneously, the following condition can be derived from Equation 3.11 and must be satisfied throughout the modulating region,

$$a \leq \frac{FWHM(\Delta)}{2} = \frac{\lambda}{2\mathcal{F}} \quad (3.13)$$

That is,

$$\mathcal{F} \leq \frac{\lambda}{2a} \quad (3.14)$$

The result of violating this condition was aptly demonstrated by the prototype F-P MSLM built, where $a \approx \lambda/4$ and $\mathcal{F} \approx 8.76$. Substituting into Equation 3.14,

$$\mathcal{F} = 8.76 \leq \frac{\lambda}{2\lambda/4} = 2 \quad (3.15)$$

Alternatively stated, $\mathcal{F} = 8.76$ ($R = 0.7$) would require a crystal with $a \leq \lambda/17.5$. Thus, high \mathcal{F} F-P MSLMs are severely limited by the maximum achievable plane-parallelism of the polished crystal surfaces by present technology. With the present technology capable of achieving a maximum $a \approx \lambda/4$, \mathcal{F} is restricted to $\mathcal{F} \leq 2$

(corresponding to $R \leq 0.25$). These limitations are not easily overcome using state-of-the-art polishing techniques, especially for the thin crystals required.

A possible alternative would be the use of a thin film modulator material deposited onto an optically flat glass substrate. This would simultaneously ensure the plane-parallelism of the modulator and improve the spatial resolution of the modulator.

A method for circumventing the plane-parallelism requirement (and enable use of the presently available crystals), has been suggested for use with visible F-P MSLMs which would employ a photocathode. In this case, the crystal parallelism deviations are initially compensated with the appropriate charge distribution before any image processing is done. This is accomplished by an optical feedback configuration where the reflection of a uniform readout light is fed back to the photocathode until the internal device nonuniformities are compensated and a stable equilibrium has been attained. This method is discussed at length in the following chapter of this thesis.

Chapter 4

Optical Feedback Configuration

The optical feedback configuration of Figure 4.1 has been suggested as a means for: (1) compensating phase variations introduced by the modulator crystal, and (2) realizing bistable operation. Phase compensation is especially important for high finesse Fabry-Perot MSLMs which would otherwise have extremely stringent crystal plane-parallelism requirements for image processing, unattainable by present polishing technology.

The system or input under investigation is placed in the input plane (at P_1) or in the feedback path (at P_3). A variable spatially uniform amplitude attenuation plate is also placed in the feedback path at P_2 to control the feedback level. The output is recorded in plane P_3 . Note that in this case the intensity of the signal fed back is proportional to the intensity of the output signal beam.

The crossed polarizers in Figure 4.1 are omitted for z-cut crystals since these are read out interferometrically. The MSLM may be operated in the closed loop (feedback) mode by opening the shutter S, and in the open loop mode by closing the shutter. The MSLM output is monitored in the plane P_3 , and is equal to some fraction of the feedback signal. A procedure for operating both a standard or

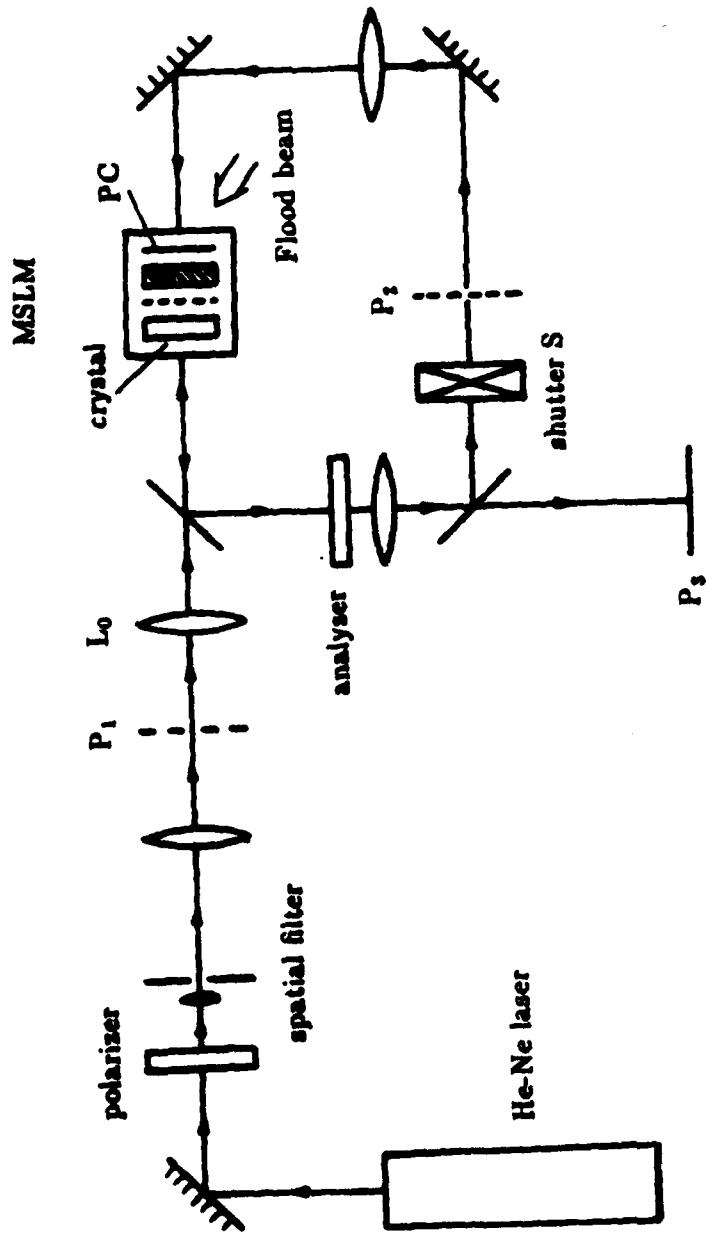


Figure 4.1: Optical feedback configuration of the MSLM

Fabry-Perot MSLM in this configuration would be:

1. compensate, closed loop
2. bias crystal, open loop
3. write nonlinear image, closed loop

Phase compensation and bistable operation are accomplished as follows. The input object is removed (uniform input), the feedback path is closed (S open) and the device is operated, preferably in the electron deposition mode, until the readout intensity in plane P_3 is uniformly zero.

In this closed loop operation, areas of non-zero transmittance of the crystal will result in feedback to the photocathode and deposition of surface charge on the corresponding areas of the crystal so as to drive the transmittance of the modulator at that region to zero. The intrinsic phase variations of the MSLM are thus compensated with the appropriate amount of surface charge and a stable equilibrium is attained.

The device should now be biased to a bright background by operation in the open loop mode (S closed) and again flooding the photocathode with light and adjusting V_s to add or remove a uniform charge density while still maintaining the phase compensation profile. The flood light is then shut off and V_s set to a large enough voltage such that the gap voltage V_g remains high enough to ensure linear operation throughout the imaging process.

The image transparency to be processed is then placed in the input plane P_1 and transferred by lens L_0 onto the crystal with the shutter S open. The resulting feedback signal is once again reimaged via the feedback path onto the photocathode. In this case, for a given fixed incident intensity on the input image,

the high transmittance regions of the input image will at first result in strong feedback and rapid deposition of sufficient surface charge on the crystal so as to drive the modulator transmittance of those points to zero (stable equilibrium) within the allowed feedback time t_w . On the other hand, low transmittance regions of the input image result in weak feedback, which will in turn result in very small amounts of additional charge deposition on the crystal. Thus, at the end of the feedback time t_w , the transmittance of the modulator remains high. The MSLM stores these charge patterns and when read out in open loop mode using a uniform source, a bistable image of the input transparency results.

The theory of the MSLM operation in this optical feedback configuration can best be modelled by considering the behaviour of a single pixel of the modulator and neglecting any transverse effects from neighbouring pixels.

The intensity of the feedback signal, $I_f(t)$ can be written as,

$$I_f(t) = I_i T(t) \quad (4.1)$$

where I_i is the incident input intensity transferred to the crystal, $T(t)$ is the particular time-dependent transmittance and t is the elapsed feedback time. $I_o(t)$, the output signal is some fraction of the feedback signal (see Figure 4.2). For a standard MSLM, the transmittance $T(t)$ (from Equations 2.1 and 2.2) is given by,

$$T(t) = \sin^2 \left(\frac{\Gamma(t)}{2} \right) = \sin^2 \left[\frac{\pi \sigma(t)}{2\lambda C} (n_e r_{r'} - n_o r_{s'}) \right] \quad (4.2)$$

where $\sigma(t)$ is the total accumulated surface charge density on the crystal at time t . In the linear mode of operation the incremental charge accumulated in the time dt is derived from Equation 2.5 and is given by

$$d\sigma(t) = \left(\frac{\eta e G}{h\nu} \right) I_f(t) dt \quad (4.3)$$

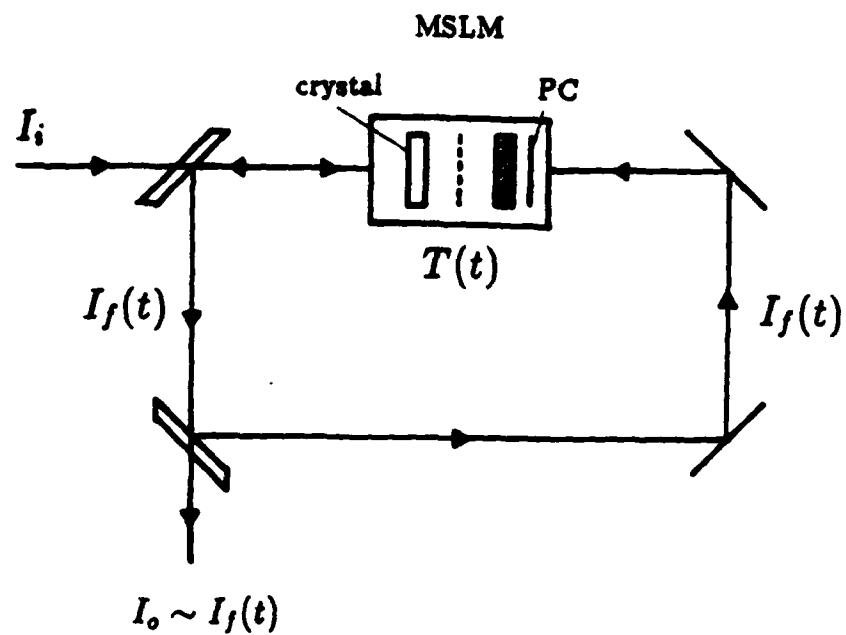


Figure 4.2: Functional block diagram of the closed-loop configuration

The total charge density integrated on the surface of the crystal at time t is therefore,

$$\sigma(t) = \int_0^t \left(\frac{\eta e G}{h\nu} \right) I_f(t') dt' + \sigma_0 \quad (4.4)$$

where σ_0 is the initial charge on the crystal. Equation 4.1 can now be rewritten as,

$$I_f(t) = I_i \sin^2 \left[B \int_0^t I_f(t') dt' + \frac{\phi_0}{2} \right] \quad (4.5)$$

where ϕ_0 is the phase retardation due to σ_0 and B is a lumped MSLM constant. For the purposes of this analysis, the lumped constant B , which acts simply as a scaling parameter, is assumed to be equal to 1. The initial conditions at $t = 0$ is,

$$I_f(t = 0) = I_i \sin^2 \left(\frac{\phi_0}{2} \right) \quad (4.6)$$

which is the reflection due to the initial bias of the crystal. Notice that at time t_f ,

$$I_f(t_f) = I_i \sin^2 \left[\int_0^{t_f} I_f(t') dt' + \frac{\phi_0}{2} \right] = 0 \quad (4.7)$$

if

$$\int_0^{t_f} I_f(t') dt' + \frac{\phi_0}{2} = p\pi; \text{ where } p = 0, \pm 1, \pm 2, \dots \quad (4.8)$$

and moreover,

$$0 \leq \int_0^{t_f} I_f(t') dt' = p\pi - \frac{\phi_0}{2} \leq \pi \quad (4.9)$$

Once the feedback signal $I_f(t)$ is driven to zero at time t_f , no further charge deposition occurs, so that,

$$I_f(t) = 0 \text{ for all } t \geq t_f, \quad (4.10)$$

and stable equilibrium is achieved. Notice that for phase compensation, the shutter S must remain open long enough to allow the entire crystal to achieve equilibrium. This behaviour is monitored in the output plane.

Alternatively, for nonlinear processing (e.g. bistable operation) feedback is allowed only long enough for areas of the input with transmittance above some threshold to be driven to equilibrium. The threshold transmittance is determined by the feedback time t_w . In the nonlinear processing case, t_w is such that,

$$\int_0^{t_w} I_f(t') dt' + \frac{\phi_0}{2} \approx \begin{cases} p\pi & \text{for high input transmittance} \\ \frac{\phi_0}{2} & \text{for low input transmittance} \end{cases} \quad (4.11)$$

Thus the modulator transmittance will be given by,

$$T(t_w) \approx \begin{cases} 0 & \text{for high input transmittance} \\ \sin^2\left(\frac{\phi_0}{2}\right) & \text{for low input transmittance} \end{cases} \quad (4.12)$$

and the desired bistable image will be attained upon readout in the open loop configuration with a uniform source.

The dynamic behaviour of the feedback signal is best analysed by numerically calculating and plotting $I_f(t)$. From the fundamental theorem of calculus, Equation 4.5 can be estimated as follows,

$$I_f(t = N \Delta t) = I_i \sin^2 \left[\sum_{n=1}^N I_f((n-1), \Delta t) \Delta t + \frac{\phi_0}{2} \right] \quad (4.13)$$

where the feedback signal is considered constant during the incremental time interval $\Delta t = t/N$; N is the number of time intervals. The initial condition is still given by Equation 4.6.

Equation 4.13 can be analysed for two cases: (1) varying the initial bias phase ϕ_0 due to the crystal nonuniformities, and (2) varying the input intensity I_i transmitted by the input transparency.

Figure 4.3a contains plots of $I_f(t)$ for multiple values of the initial phase ϕ_0 with constant input intensity I_i . The transmission will decay to zero faster than the characteristic sine-squared curve and equilibrate. Notice that in cases where $\phi_0/2$ occurs before a maximum of modulator transmission, the feedback signal will actually increase first before decaying to zero.

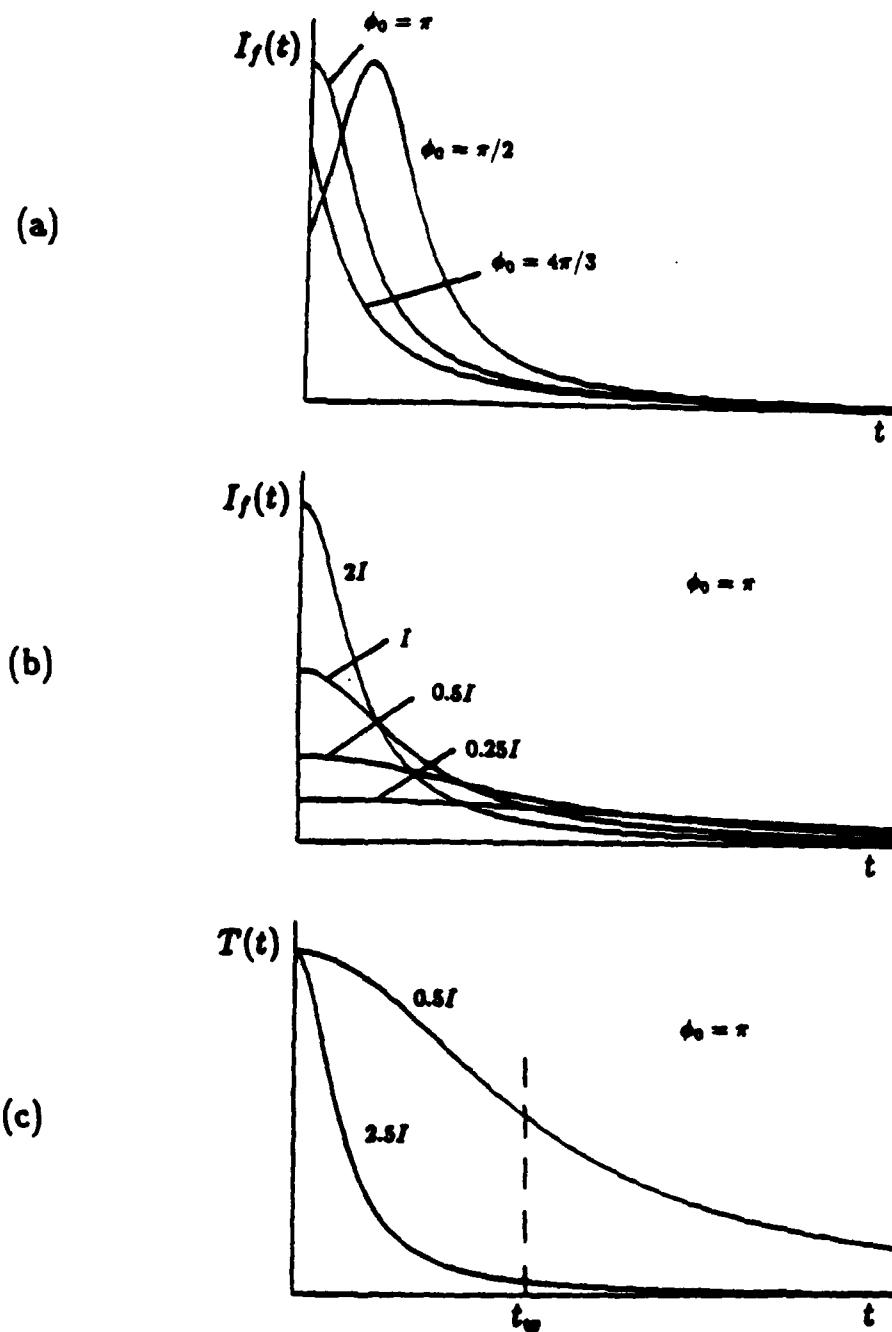


Figure 4.3: Output intensity $I_f(t)$ of a standard MSLM in the closed-loop configuration: a) for constant input intensity I_i , with initial phase ϕ_0 as a parameter, b) for constant initial phase $\phi_0 = \pi$ and with input intensity I_i as a parameter, and c) corresponding modulator transmission $T(t)$

The effect of different input intensities for the same initial phase $\phi_0 = \pi$ on $I_f(t)$ is shown in Figure 4.3b, and the effect on the modulator transmittance $T(t)$ is shown in Figure 4.3c. In this case, ϕ_0 is set at π radians and the relative values of the input intensities varied. Notice that $I_f(t)$ (and thus, $T(t)$) decays to zero more rapidly for higher input intensity. This supports the discussion on the use of this feedback configuration for nonlinear applications. At some time t_w , indicated in Figure 4.3c regions with high I_i are driven to zero transmittance, while those with low I_i leading to very little additional charge deposition, maintain relatively high modulator transmittance.

A similar analysis has been done for the dynamic behaviour of the Fabry-Perot MSLM. For this version of the device, the transmittance $T(t)$ is derived from Equations 3.2 and 4.1 rewritten as

$$I_f(t) = I_i \frac{4R \sin^2 [B' \int_0^t I_f(t') dt' + \frac{\pi}{2}]}{(1 - R)^2 + 4R \sin^2 [B' \int_0^t I_f(t') dt' + \frac{\pi}{2}]} \quad (4.14)$$

where ϕ_0 is again the initial bias phase and B' is a lumped device constant. Once again, B' is a scaling parameter and is assumed equal to 1. The initial condition is,

$$I_f(0) = \frac{I_i 4R \sin^2 \left(\frac{\pi}{2}\right)}{(1 - R)^2 + 4R \sin^2 \left(\frac{\pi}{2}\right)} \quad (4.15)$$

The feedback signal will again equilibrate at zero when $t \geq t_f$, where t_f is still as defined in Equation 4.8. The Fabry-Perot device will operate in much the same manner as the standard device.

The high finesse Fabry-Perot device will compensate for crystal deviations from plane-parallelism if sufficient feedback time is allowed, and can achieve nonlinear operation when the feedback time is appropriately limited. Equation 4.14 was analysed numerically and plots were made for: 1) $I_f(t)$ with constant I_i , and

different ϕ_0 (shown in Figure 4.4a), and 2) $I_f(t)$ and the corresponding $T(t)$ with constant $\phi_0 = \pi$ and different relative I_i (shown in Figure 4.4b and c).

Notice that the intrinsic nonlinear characteristics of a high finesse Fabry-Perot crystal ($R = 0.7$) result in similar nonlinear behaviour of the feedback signal and the modulator transmittance. In nonlinear operation, the threshold input transmittance can be precisely determined by t_w , and the resulting bistable image is significantly improved over that obtained with a standard device.

The disadvantage of using this type of feedback configuration to compensate for crystal deviations from plane-parallelism by charge deposition, is that there is a trade-off in dynamic operating range. Experimentally the safe dynamic operating range of an MSLM device is about 4kV; this limitation is imposed by arcing between components and stresses due to high electrostatic potentials. Consider a standard device with $V_{zR} = 1250V$; if the MSLM to be compensated has a crystal with plane-parallelism factor (see Equation 3.12), $a = \lambda/4$, then the crystal will require about $V_{zR}/2 = 625V$ equivalent charge compensation in some areas. Thus, about 15% of the available operation range is utilised in compensation, leaving over 3kV to work with; not unreasonable for a crystal with $V_{zR} = 1250$ (full modulation can be realised for a grid voltage of 1 – 2kV).

However, the compensation trade-off is much more severe in the case of the Fabry-Perot. Consider a device with a z-cut LiNbO₃ crystal, $V_z = 3100V$; once again assuming $a = \lambda/4$, the crystal will require $V_z/2 = 1550V$ equivalent charge compensation. This is about 40% of the available operating range, leaving only about 2450V with which to work. Although this range is greater than the voltage required for a light to dark transition (600 – 800V; $R = 0.7$), if the grid voltage is 1 – 2kV, no allowance would be made for the fact that the precise thickness of the

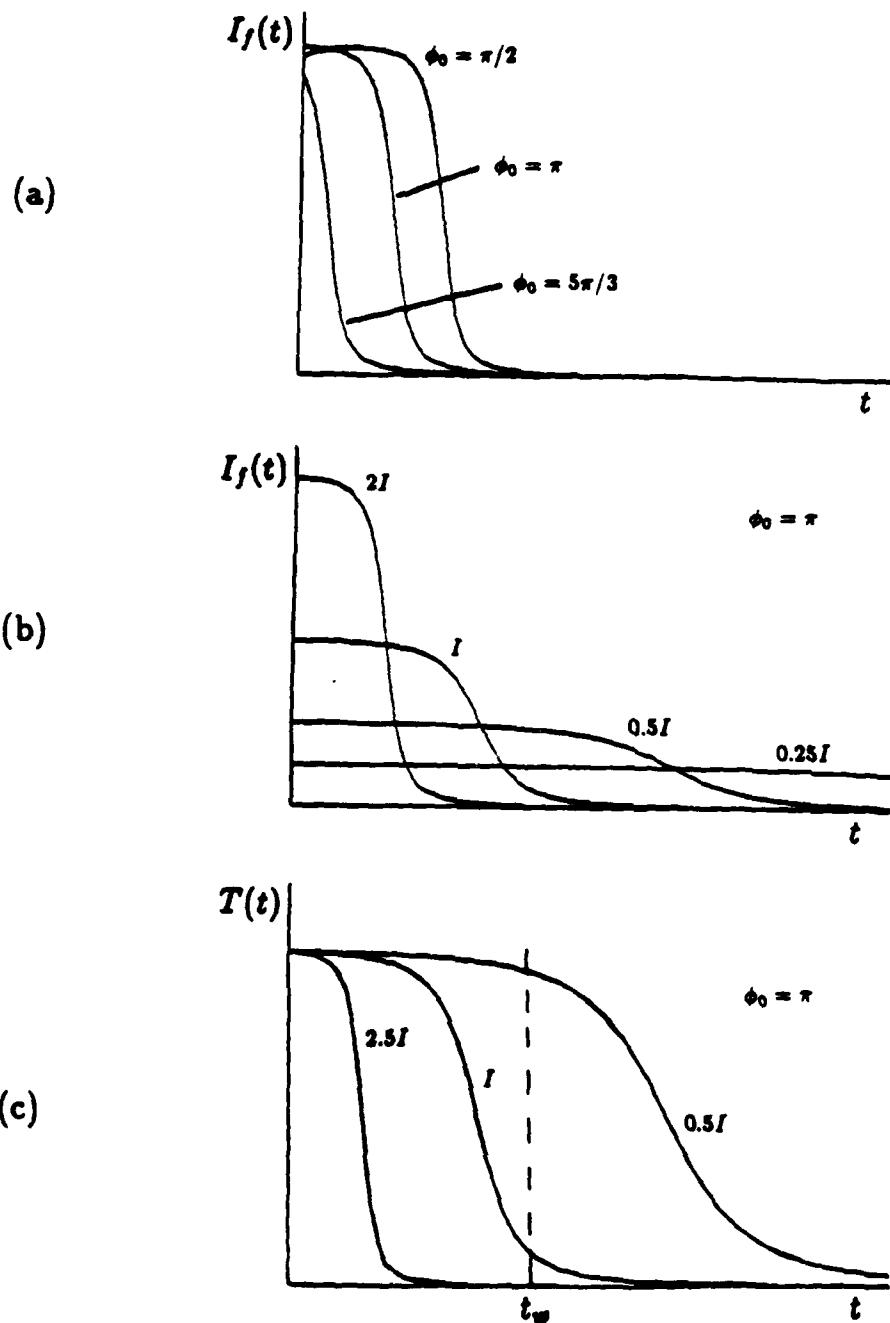


Figure 4.4: Output intensity $I_f(t)$ of a Fabry-Perot MSLM in the closed-loop configuration: a) for constant write intensity I_i with initial phase ϕ_0 as a parameter, b) for constant initial phase $\phi_0 = \pi$ and with write intensity I_i as a parameter, and c) corresponding modulator transmission $T(t)$

crystal can be as much as $(n + \frac{1}{2})\lambda$, where n is an integer and λ is the wavelength of the readout light. Thus, the crystal would be biased in the neighborhood of the central flat region of the Fabry-Perot characteristic. This additional fraction of a wavelength must be accounted for experimentally by properly biasing the operating voltages and in the worst case can require as much as $V_s = 3100V$.

The ability to compensate internal device phase variations is critical for imaging with a Fabry-Perot MSLM using currently available crystals. However, the amount of phase compensation that can be tolerated while leaving sufficient operating range, is limited by the precise crystal thickness deviation from an integer multiple of λ , and the voltage limitations imposed on the device.

Chapter 5

Variable Gamma Operation

The goal of the work described in this chapter was to convert the inherent periodic characteristic of the MSLM (Figures 2.6 and 3.1) to a switching characteristic useful in nonlinear processing by choice of operating mode. The next Section 5.1 contains a description of the conventional gamma characteristic used for describing the switching characteristics of photographic emulsions. This is followed by a discussion of the desired switching behaviour and the parameters which characterise the response of the MSLM. The following sections describe the operating modes which were used to convert the periodic MSLM characteristic to a switching characteristic and an analysis of the modulator transmittance with respect to the individual operating parameters.

The techniques for achieving the desired MSLM characteristics are:

- Electron deposition/grid stabilized mode
- Electron depletion/adjustable ramp rate, fixed terminal ramp voltage (*Real-Time Hardclipped Thresholding Mode*)

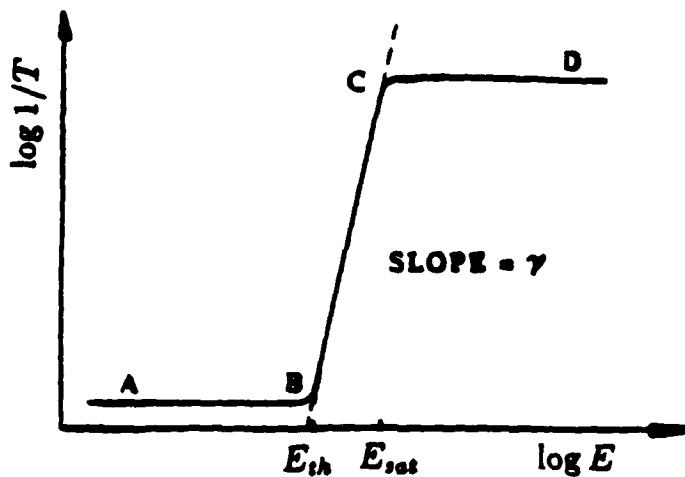


Figure 5.1: Conventional gamma characteristic for photographic film

5.1 Switching Characteristics

5.1.1 Conventional Gamma Characteristic of Photographic Film

Recall that the gamma characteristic, defined for conventional photographic emulsions, is a plot of $\log 1/T$ vs. $\log E$; where T is the readout intensity transmission and E is the corresponding input exposure (see Figure 5.1). The input exposure is the product of the incident exposure I_o and the exposure time t_o . For a negative gamma characteristic the readout light remains in the *ON* state (high T) for input exposures below threshold E_{th} , and is *OFF* (low T) for exposures above saturation E_{sat} . The gamma γ of the device is the slope of the linear region of the curve,

and is given by,

$$\gamma = \frac{\log(T_{th}/T_{sat})}{\log(E_{sat}/E_{th})} \quad (5.1)$$

The gamma characteristic is more commonly referred to as the *Hurter-Drifford (H and D)* curve in the photographic world, and is the most widely used description of the photosensitive properties of photographic film. Typical values of gamma for *low contrast* film are 1 or less, and for *high contrast* films, 2 or 3 [11].

The gamma characteristic for film is uniquely defined in terms of the threshold E_{th} and saturation E_{sat} ; that is, there is a single unique gamma associated with a given emulsion. The logarithmic nature of the $T - E$ relationship allows the value of γ to be conveniently determined with respect to either I_w or t_w ; the log decouples the gamma dependence on the product $\log(I_w t_w)$, to the sum of $\log(I_w)$ and $\log(t_w)$. Thus, γ may be measured by either varying I_w or t_w . Note that while it may be possible to vary the threshold and saturation exposure through the development process, and thus obtain a different value for γ , the switching nature of the gamma characteristic is preserved. A positive gamma characteristic requires two recording steps if one starts with a negative film.

5.1.2 Desired Switching Characteristic of an MSLM

The *ABC* region of the standard MSLM characteristic and that of the Fabry-Perot version (Figure 5.2) closely approximates the corresponding region of the gamma characteristic of film (Figure 5.1). However, the periodic nature of the MSLM readout characteristics precludes the unique definition of threshold, saturation and γ for the MSLM. By properly operating the device, the desired switching characteristic shown in Figure 5.3 can be attained; in this case, the device is forced to follow the curve *ABCD* instead of its normal characteristic by choice of

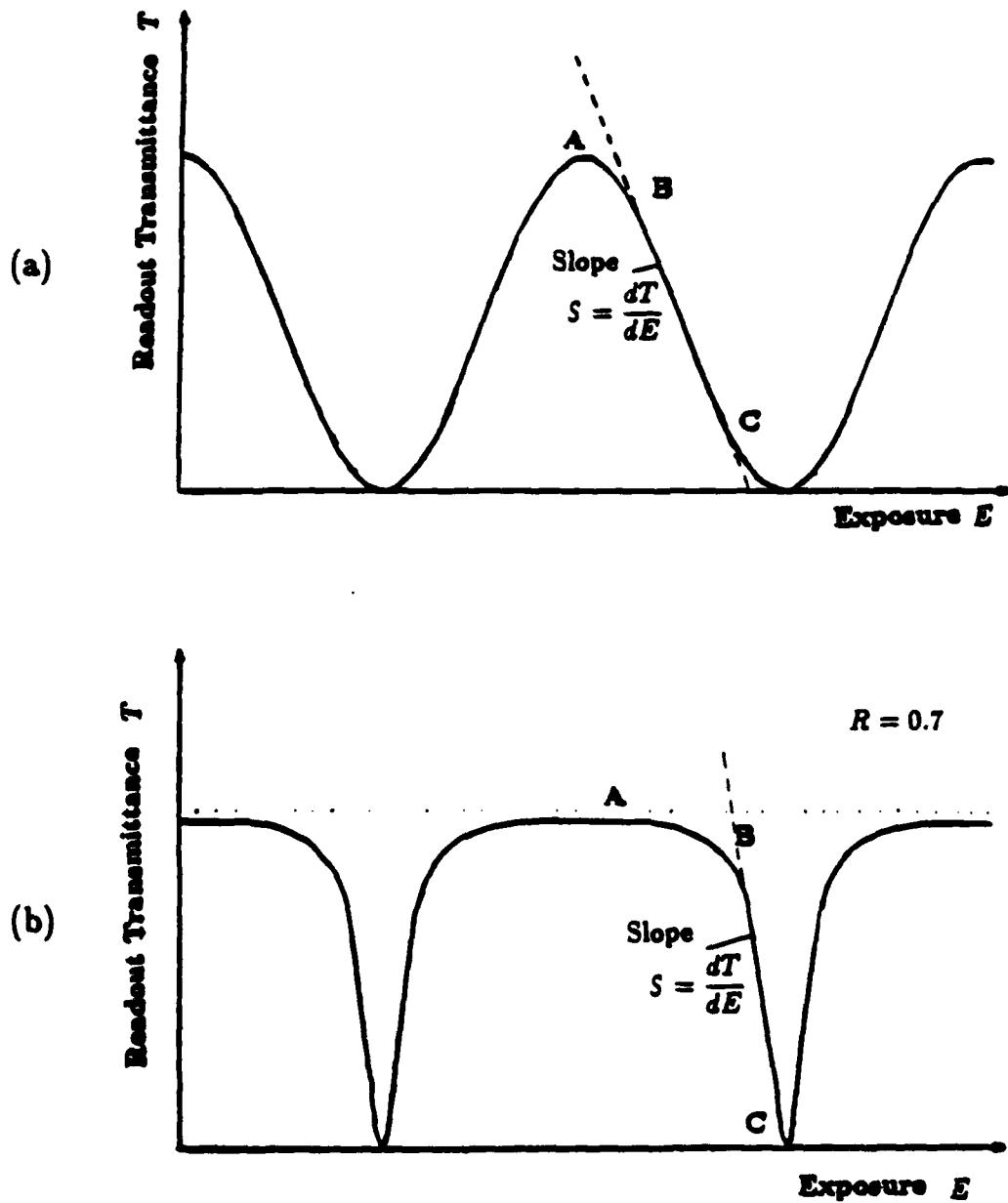


Figure 5.2: a) Polarization readout transmission characteristic of an oblique-cut LiNbO_3 crystal and b) interferometric readout transmission characteristic of a Fabry-Perot crystal with $R = 0.7$

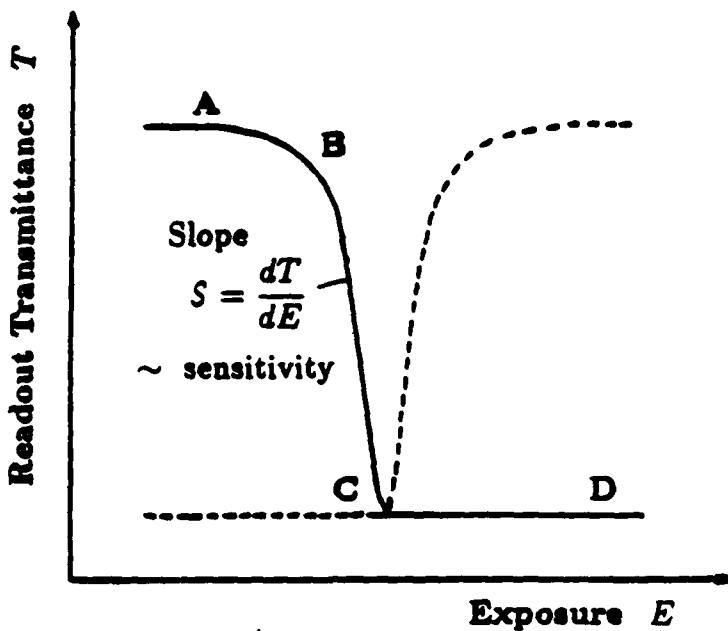


Figure 5.3: Desired MSLM switching characteristic

operating mode.

According to Equation 5.1 the value of γ is dependent on the ratio E_{sat}/E_{th} of the quasilinear region in the log - log plane; this ratio will be single-valued for a given photographic emulsion. However, the MSLM initial bias point A may be set anywhere along the characteristic curve by initially depositing the appropriate surface charge σ_A on the crystal. For a given bias point, there will be an associated threshold exposure E_B and saturation exposure E_C . The ratio of the saturation to threshold exposure E_C/E_B is therefore bias dependent, and thus, γ will also be bias dependent. Notice that although the value obtained for γ in the $\log 1/T$ vs. $\log E$ plane changes with the initial bias, the slope of the linear transition region of the intrinsic MSLM characteristic in the $T - E$ plane is uniquely defined. This dilemma is especially apparent in the case where the

MSLM bias A is very close to the threshold at point B . In this case, $E_B \approx 0$ and γ becomes undefined; however, the output response of the device may still be predicted using the readout switching characteristic.

These observations indicate that the definition of the conventional gamma characteristic for photographic emulsions is inappropriate for describing the switching behaviour of the MSLM. Instead, the MSLM is better characterised by the respective readout transmission characteristic in a linear $T - E$ plot.

Thus, the slope of the transition BC region in $T - E$ plots indicated in Figures 5.2 and 5.3 will be used to describe the MSLM switching behaviour. The initial bias of the MSLM at point A serves as the modulator input exposure reference point and points B and C correspond to the threshold and saturation, respectively. For convenience E_B and E_C are defined at 90% and 10% transmission of the modulator readout intensity. The slope S in the linear region of the $T - E$ curve for the modulator is given by,

$$S = \frac{dT}{dE} = \frac{T_C - T_B}{E_C - E_B} \quad (5.2)$$

The definition of the slope S is now independent of the initial bias A . Both negative and positive values of S are possible.

Notice that the exposure dependence E of the slope can no longer be decoupled with respect to I_w and t_w as in the case of photographic film; instead, both quantities must now be measured to evaluate S . The slope S has units of transmittance times inverse exposure, and can be considered as a measure of the device response or sensitivity to the input exposure. Alternatively stated, S measures the change in the device transmission per unit input exposure. A high value of S implies a *high sensitivity* device (high device response to exposure), while a *low* value of S implies *low sensitivity* device (low device response to exposure). A high

S device is therefore desirable in nonlinear operations requiring rapid switching in the output transmission.

5.2 Electron deposition/Grid stabilized Mode

Both the standard and the Fabry-Perot MSLM may be operated in the electron deposition/grid stabilized mode. This is accomplished by first biasing the crystal with a bright background at point A , setting the grid voltage, V_{grid} to point C (the voltage required to deposit the required surface charge to bias the crystal at point C). The image is then written by exposing the photocathode to the write light I_w for a specified write time t_w . Thus, areas of input exposure below E_B will be barely recorded, those above E_C will saturate at C since the grid voltage prevents further charge deposition, and exposures in between are determined by the slope S .

In the electron deposition/grid stabilized mode of operation, the value of S obtained is intrinsic to the particular MSLM device type. The device parameters which determine the intrinsic modulator slopes for both the standard and F-P MSLMs are examined as follows.

The sine-squared characteristic of a standard MSLM is approximately linear for transmission between 10% and 90%. The slope of this linear region may be approximated to within 15% error by the value of the derivative at half-maximum ($\Gamma = (n + \frac{1}{2})\pi$, n integer) by the following method (neglecting sign for now).

Consider,

$$\left. \frac{dT}{d\Gamma} \right|_{\Gamma=(n+\frac{1}{2})\pi} = \frac{1}{2} \quad (5.3)$$

From Equations 2.1, 2.3 and 2.5, obtain,

$$\Gamma = \frac{\pi\sigma(E)}{CV_{sR}} = \left[\frac{\pi}{CV_{sR}} \frac{\eta eG}{h\nu} \right] E \quad (5.4)$$

where s refers to $[\cdot]$ evaluated using the appropriate standard MSLM parameters.

The slope S can now be estimated,

$$S_s \approx \frac{dT}{dE} = \frac{1}{2} \left[\frac{\pi\eta eG}{CV_{sR}h\nu} \right]. \quad (5.5)$$

The Fabry-Perot version of the MSLM is much more difficult to analyse using the above technique. Instead, the corners of the Fabry-Perot curve are assumed to be sharp; once again neglecting the sign, Equations 3.6 and 3.7 lead to the following,

$$\frac{dT}{d\Gamma} \approx \frac{1}{FWHM(\phi)} = \frac{\mathcal{F}}{2\pi} \quad (5.6)$$

From Equations 2.5, 3.1 and 3.5, obtain,

$$\Gamma = \frac{\pi\sigma(E)}{CV_s} = \left[\frac{\pi}{CV_s} \frac{\eta eG}{h\nu} \right]_{F-P} E \quad (5.7)$$

where $F\text{-}P$ refers to $[\cdot]$ evaluated using the Fabry-Perot device parameters. Now, the slope S can be estimated,

$$S_{F-P} \approx \frac{dT}{dE} = \frac{\mathcal{F}}{2\pi} \left[\frac{\pi\eta eG}{CV_s h\nu} \right]_{F-P} \quad (5.8)$$

In both cases the photocathode quantum efficiency η and MCP gain G increase the device response, while the crystal capacitance C and halfwave voltage V_s decrease the response as expected. However, in the case of the F-P device, there is an additional factor \mathcal{F}/π contributing to S . This makes a high finesse F-P modulator significantly more desirable in nonlinear operations. Consider a standard and a F-P device (mirror reflectivity $R = 0.7$); then, if $[\cdot]_s \approx [\cdot]_{F-P}$, the ratio of the device slopes will be about,

$$\frac{S_{F-P}}{S_s} \approx \frac{\mathcal{F}}{\pi} = 2.8 \quad (5.9)$$

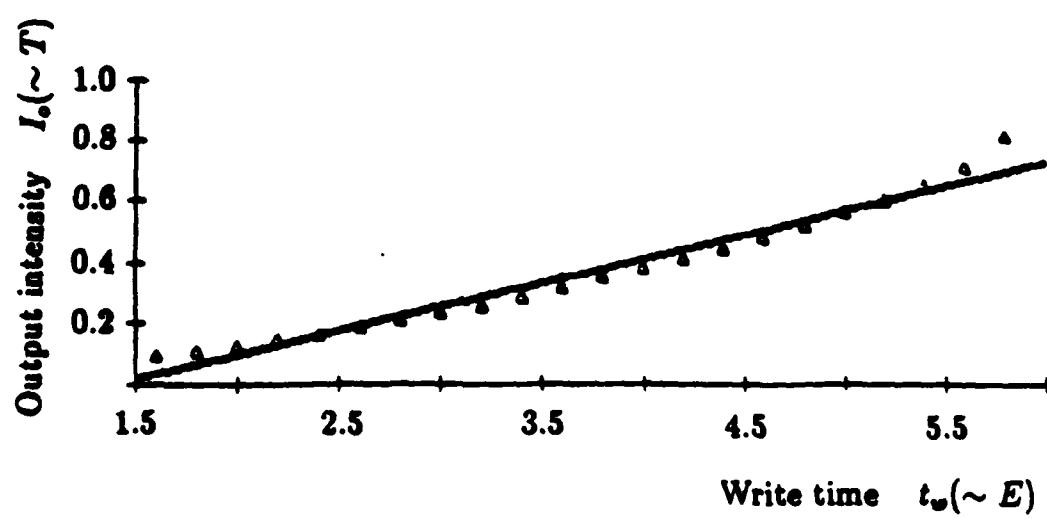
The F-P version sensitivity is a factor of 2.8 greater than the standard device. Thus, full light-to-dark transitions can be achieved with much less input exposure.

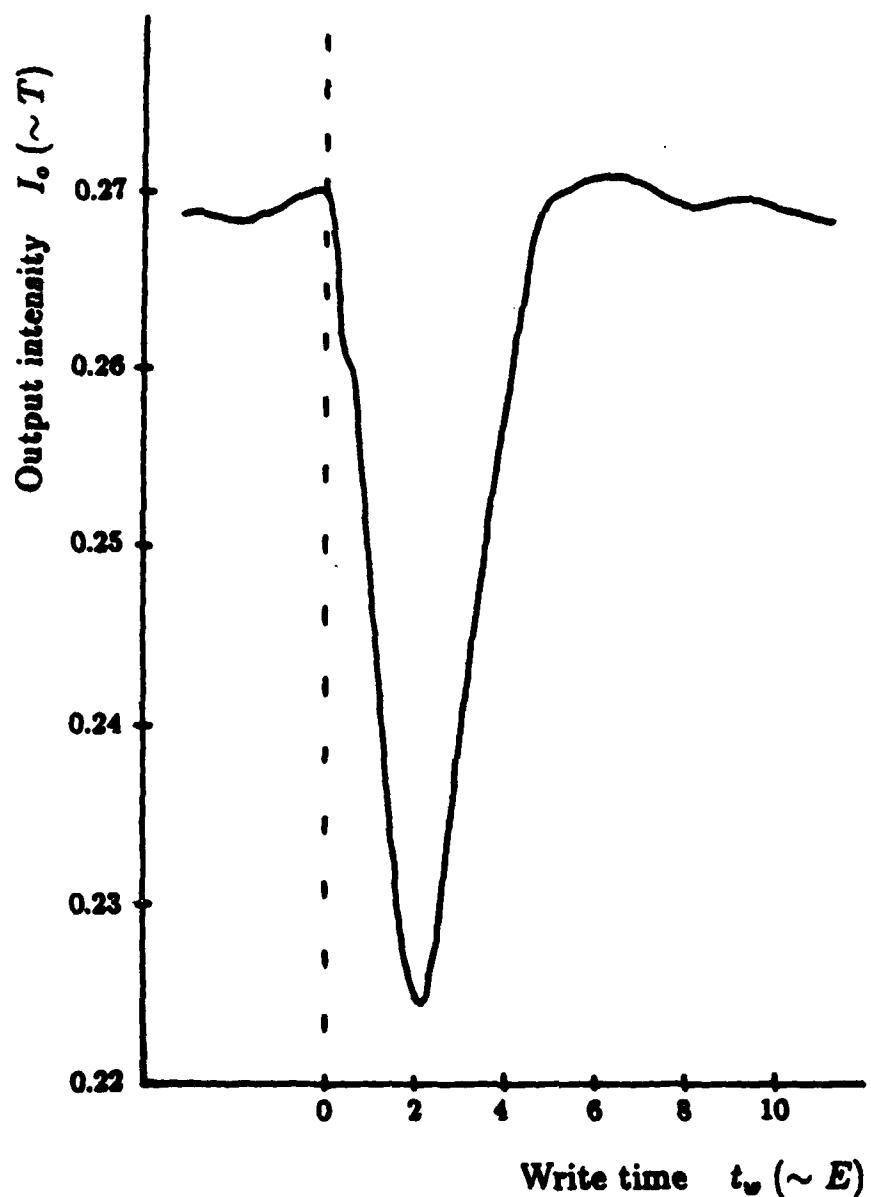
As mentioned before, the slope obtained in this operating mode is intrinsic to the device and is dependent on the appropriate device parameters. This slope is independent of the incident intensity and exposure time used. This is especially significant in the following section which describes the hardclipped thresholding mode which modifies the intrinsic behaviour of the device and is characterised with respect to the write intensity and the write time.

Measurements of the absolute input intensity level on the photocathode requiring carefully calibrated detectors were not conducted, so that the absolute numerical values for S were not evaluated. However, calibrated neutral density filters were used to attenuate the incident intensity in a controlled manner, and thus, enabled analysis of the experimental data with respect to relative intensity levels.

Figure 5.4 shows the data obtained using the standard Hamamatsu MSLM. The plot of the raw data I_{detout} vs. t_w , where I_{detout} is the detector measurement corresponding to the intensity I_o in the output plane. While the write intensity level I_w was not measured, because it remained constant throughout the measurement, the elapsed write time was proportional to the input exposure ($E = I_w t_w$). The best fit to the data by linear regression is also shown as a solid line in the figure.

The F-P MSLM intrinsic characteristic was also measured, but in this case a chart recorder was used to plot the amplified detector signal directly. Figure 5.5 shows a plot of I_{detout} vs. t_w for the prototype F-P device that was built. Once again the input write UV intensity was constant but unmeasured. According to

Figure 5.4: I_o vs. t_w for standard Hamamatsu MSLM

Figure 5.5: I_o vs. t_w for prototype Fabry-Perot MSLM

the time scale for these typical write intensities used in recording the intrinsic device S , the F-P is probably more sensitive than the standard device. Absolute values of the slope S would require measurement of the average write intensity at the appropriate wavelength of write light.

The quasilinear behaviour of the MSLM is apparent in both cases and the validity of using a linear model in the transition BC region of Figure 5.3 is substantiated. While this mode of operation is capable of achieving the desired switching characteristic, it is not the preferred mode of operation. This is because the saturation point C is not well-defined experimentally and proximity focusing is lost when areas of the crystal saturate near the grid voltage. This loss of proximity focusing results in degradation of the images that are recorded. The second, more preferred technique is described below.

5.3 Real-Time Hardclipped Thresholding Mode

Both the standard and F-P MSLMs can be operated in the hardclipped thresholding mode to vary their intrinsic characteristics. To operate the MSLM with a negative slope characteristic, the device is first biased with a bright background at point A in Figure 5.3. The grid voltage V_{grid} is set at point C , then with the optical image incident on the photocathode, the crystal voltage V_i is ramped downward from point A at some pre-selected rate \dot{V}_i , to a terminal voltage which is usually selected to be the grid voltage. The rate \dot{V}_i establishes both the threshold exposure E_B (corresponding to the initial bias A) and the slope S . All exposures below E_B will be barely recorded and remain in the ON state; those above E_C will saturate at the grid voltage (OFF state); and the exposures in between will be determined by S . Remember that the output intensity is proportional to the

device transmittance T .

Experimental results were obtained using the standard Hamamatsu MSLM. As mentioned previously, due to the voltage limitations for this particular device, slopes were measured in an operating region around the $\sigma_{\pi/4}$ point.

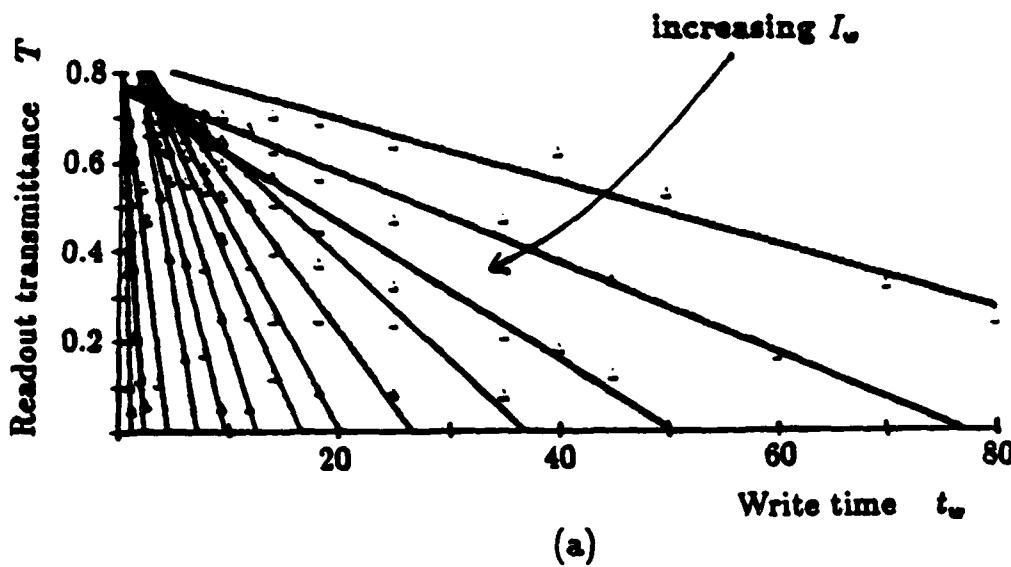
Variable-gamma and variable-threshold operation was achieved by varying the ramp rate (that is, the write time t_w from A to C), and the relative write intensity I_w . The write intensity was varied using a uniform write light attenuated by calibrated neutral density filters to obtain different values for I_w on the photocathode. Unlike photographic film, the results demonstrated that both the threshold exposure and the sensitivity (S) of the device depended on the specific values of t_w and I_w during the write operation.

Figure 5.6a is a graph of T vs. t_w with I_w as a parameter, where T is the readout transmittance, t_w is the write time and I_w is the write light intensity. Each plot was made with a constant I_w of arbitrary units. Notice that the slope of these plots dT/dt_w increases and the plots shift left for increasing I_w . This indicates that the threshold decreases when the I_w is increased. Considering only the transition region of the switching characteristic, the slope $S(I_w)$ in the $T - E$ plane is given by,

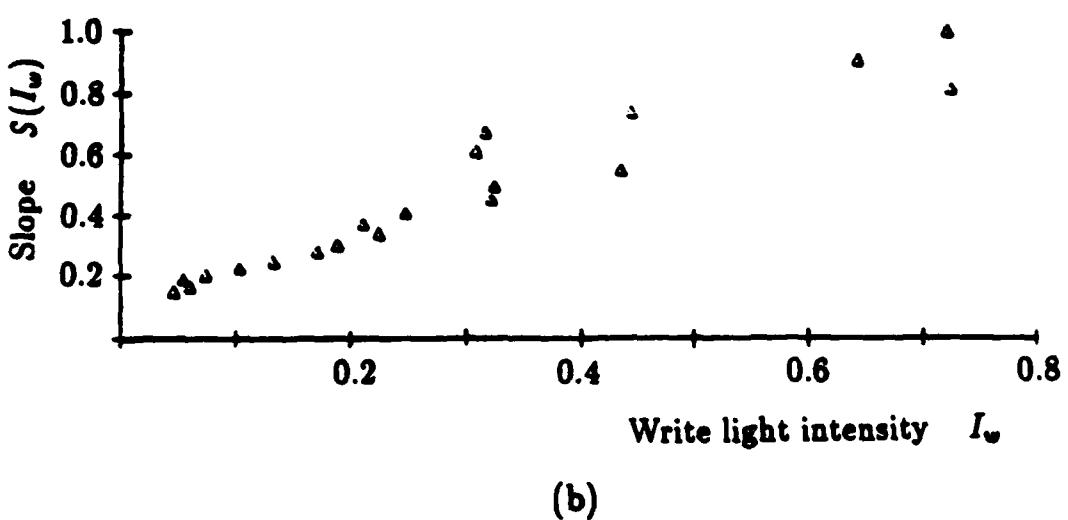
$$S(I_w) = \frac{dT}{dE}(I_w) = \frac{1}{I_w} \frac{dT}{dt_w} \quad (5.10)$$

and is plotted in Figure 5.6b as a function of I_w .

Figure 5.7a is a graph of T vs. I_w , with t_w as a parameter. Each plot was made with a constant t_w and intensity of arbitrary units. Once again the slope dT/dI_w increases with increasing t_w (decreasing ramp rate V_r), but in this case the threshold exposure is not significantly changed. Considering only the transition

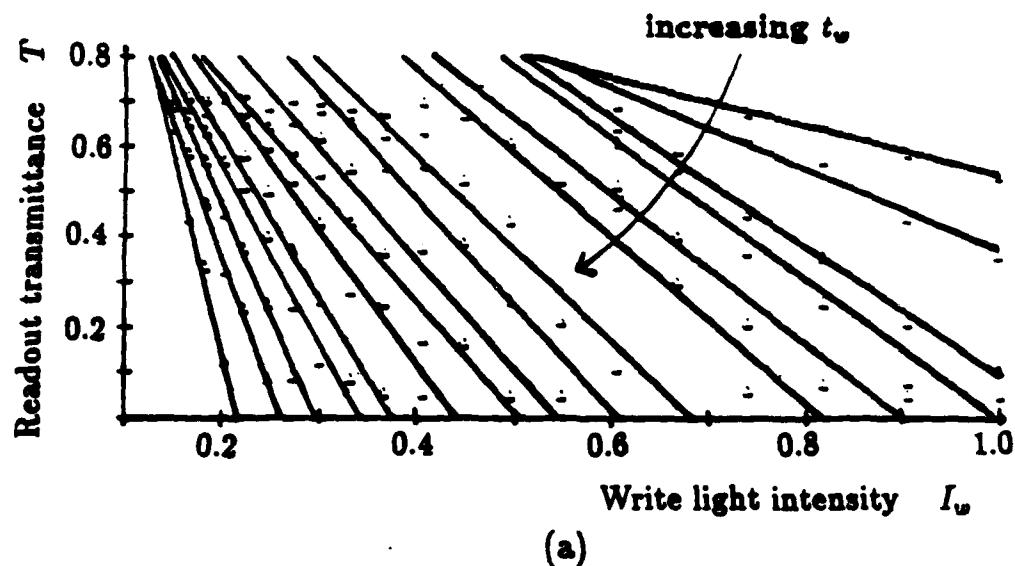


(a)

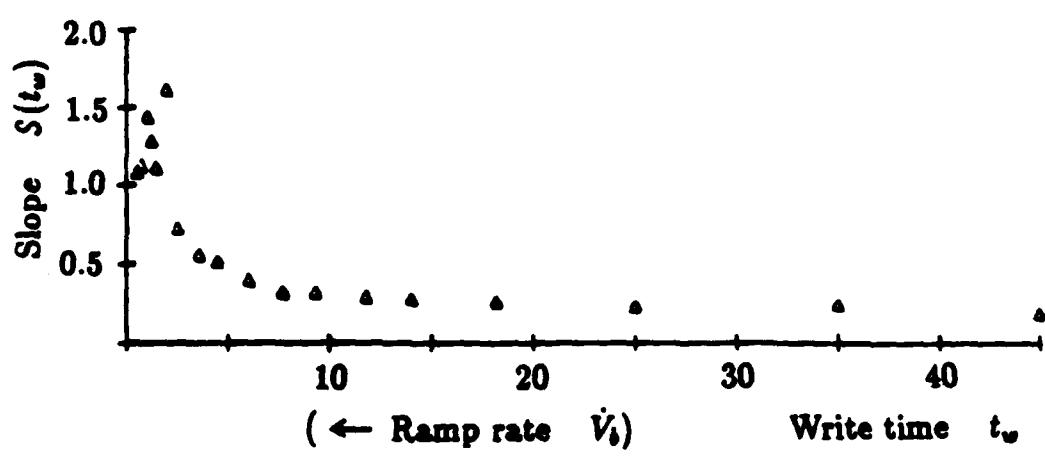


(b)

Figure 5.6: Hardclipped thresholding, with I_w as a parameter: a) T vs. t_w and, b) $S(I_w)$ vs. I_w



(a)



(b)

Figure 5.7: Hardclipped thresholding, with t_w as a parameter: a) T vs. I_w and, b) $S(t_w)$ vs. t_w

region, the slope $S(t_w)$ in the $T - E$ plane is given by,

$$S(t_w) = \frac{dT}{dE}(t_w) = \frac{1}{t_w} \frac{dT}{dI_w} \quad (5.11)$$

and is plotted in Figure 5.7b as a function of t_w .

Notice that in Figure 5.6b the sensitivity S increases with increasing write intensity I_w . This can be understood by considering the following. If the write light intensity were equal to zero then one would expect no response from the device, or, equivalently stated, the slope S would be zero. As the light intensity was increased, the device sensitivity would also increase, as indicated in the graph. One would expect at some point, further increasing of the write light intensity would result in a decrease in the derivative of the slope; since the device is incapable of responding to an infinite range in input intensity, and eventually saturation of S is approached. At saturation, the MCP is emitting the maximum number of electrons, independent of any further increase in the write light intensity, so that the additional light intensity above the MCP saturation is essentially wasted, and the sensitivity of the device remains at some maximum value.

Similar observations may be made with respect to the results shown in Figure 5.7b. In this case, the sensitivity S decreases with increasing write time t_w (corresponding to decreasing ramp rate V_s). This is a direct result of the efficiency with which the secondary emission electrons are collected by the grid. For the case when the write time is short (fast ramp rate), electrons are rapidly removed from the crystal and collected by the grid, so that the electron collection process is relatively efficient, and the device sensitivity high. As the write time increases (decreased ramp rate), the collection of electrons by the grid becomes more inefficient and is limited by the ramp rate of the voltage, resulting in a decrease in the effective device response. That is, the effective sensitivity of the device is

not limited by the write light intensity, but by the rate at which the electrons are being removed from the surface of the crystal (low V_b).

Figure 5.8 shows the results after thresholding a gray scale image of an M.I.T. student with increasing write times. Higher gamma (more nonlinear) operation would have been achieved for operation in the neighborhood of $\sigma_{\pi/2}$.

The F-P MSLM could also be operated in the hardclipped thresholding mode to achieve even higher sensitivity than the intrinsic device response. However, there is no need to operate the F-P in such a manner if the finesse is made sufficiently high to attain the desired switching characteristics. Naturally, according to the limitations discussed in Chapter 3, there is some limit to the allowed β imposed by the crystal plane-parallelism.

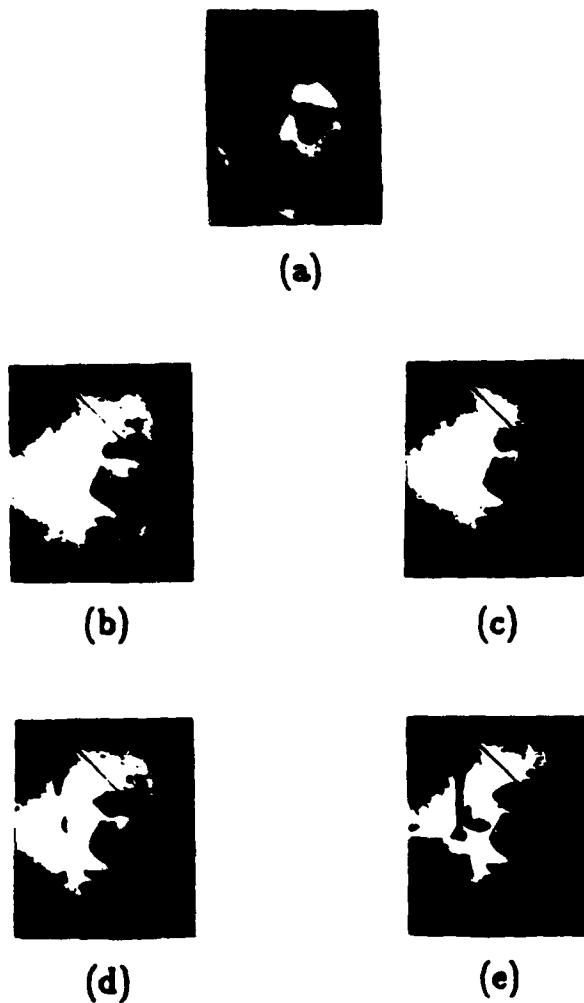


Figure 5.8: Results after hardclipped thresholding of a gray scale image of an M.I.T. student: a) gray scale image, b) - e) increasing write times

Chapter 6

Optical Information Processing

6.1 Applications Discussion

The versatility offered by a high-resolution, optically-addressed, variable-gamma spatial light modulator makes it suitable for a vast number of optical information processing applications [12,13].

At present, there exists a need to replace photographic film as the basic component of optical processing and filtering systems. If optical computers are to achieve their full potential for high speed parallel data throughput, then real-time reusable devices must become available. Such spatial light modulators are necessary in three essential areas:

- in the input data plane for amplitude or phase modulation of a coherent wavefront
- in the Fourier transform filter plane of an optical system for 2-D optical filtering
- in the output data plane for recording output of the processing system

The positive/negative variable-gamma capabilities further enhance the advantages of the MSLM for general purpose, real-time linear and nonlinear (bistable) optical processing. This section briefly reviews a few such applications. The particular application of the MSLM in real-time halftone screen processing is discussed in the following section.

Operation in a linear mode would allow the device to be utilized in most conventional linear processing. These applications include pattern recognition, Fourier plane filtering, and real-time holography. As an example, it is possible to use the MSLM in the input data plane of a coherent optical processor for incoherent-to-coherent image conversion, or in the Fourier plane as a 2-D adjustable optical filter in pattern recognition, and in the output plane as a storage medium. This stored image may then be used as the input to another optical system, since the MSLM is cascadable; this feature also allows utilization in feedback processing systems. All of these operations can be performed in real-time and preserve the inherent advantages of optical systems. These simple, yet powerful operations are basic building blocks in conventional linear processing.

Bistable modes of operation are particularly attractive for the realization of a large number of 2-D nonlinear processing operations.

A high-gamma MSLM can be used directly as a nonlinear device to demonstrate applications such as intensity level slicing, logic operations, bistability, A-D conversion and thresholding. These operations are performed in both open loop and closed loop (feedback) configurations and have been discussed in this thesis and elsewhere [10]; such operations are important for both image processing and development of a futuristic optical digital computer.

There exist several coding and modulation techniques that are essential for

the implementation of many nonlinear filtering operations by means of coherent optical systems. The most commonly used of these are the halftone screen process and theta modulation [13]. In nonlinear processing, operations such as logarithmic transformation, exponentiation, level slicing, multiple image isophote generation, subtraction and multiple image storage can be implemented.

This thesis examines the application of a standard MSLM operated in the hard-clipped thresholding mode to implement halftone screen processing and achieve a wide range of possible real-time nonlinear image and signal processing applications [10] such as histogram measurement, density slicing, density-based artificial stereo and equidensitometry (isophote production).

6.2 Halftone Screen Process

The objective of the halftone screen process is to perform intensity to spatial frequency conversion on the input image so that subsequent Fourier plane filtering results in the desired nonlinear operation [14,15,16,17,18]. This modulation step is also called *spatial pulse-frequency modulation*. Modulation is accomplished through the use of a contact screen. This contact screen is composed of an array of cells with one- and two-dimensional frequency and transmittance variations within each cell, $T_m(x, y)$. A continuous-tone input image $E(x)$ is sandwiched together with the contact screen and the combined transparency is imaged onto a high-gamma (ideally infinite) storage medium, such as the nonlinearly operated MSLMs discussed in this thesis. The resulting halftone image consists of a periodic array of area-modulated cells (see Figure 6.1 [14]).

Since the modulation step is essentially a sampling process (the intensity levels are encoded as binary pulses of varying area), the frequency of the contact screen

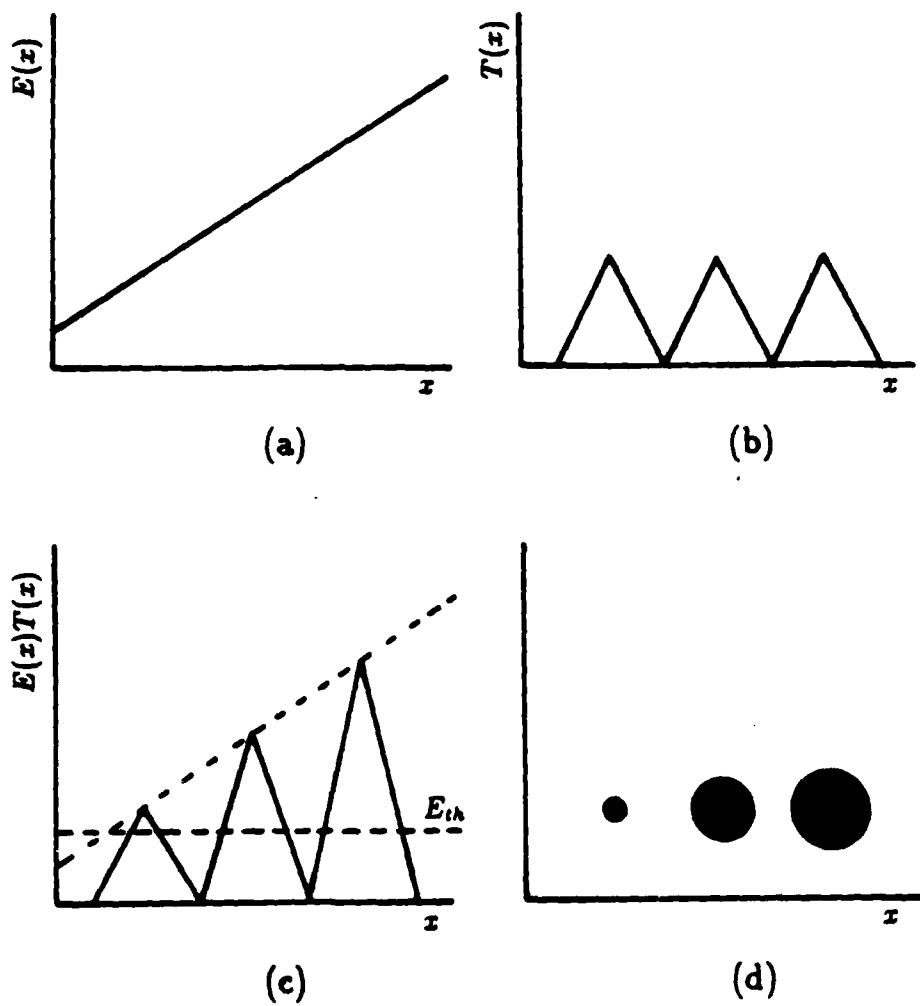


Figure 6.1: Halftone screen processing: basic concept [14], a) linearly varying input $E(x)$, b) contact screen transmission profile $T(x)$, c) incident exposure on high-gamma storage medium $E(x)T(x)$ with threshold E_{th} , d) Area modulated cells

must be sufficiently high so that within a unit cell, the transmittance of the input transparency is essentially constant.

The modulated halftone image is then fed into the input plane of a coherent optical processor which separates the various density levels into specific spatial frequency domains within the Fourier plane. The amount of diffraction is determined by the cell separation in the halftone image, and is thus a function of the MSLM threshold level. With proper design of the screen, spatial filtering in the Fourier plane can be effectively utilised to achieve a wide range of nonlinear processing operations [16,17,18].

The transmittance profile of the contact screen that was actually used to perform the halftone screen processing was provided by H. K. Liu of the Jet Propulsion Lab, and is shown in Figure 6.2a. Consider a y -slit extended in the x -direction, extracted from the center of the cell with the intensity profile of Figure 6.2b. The unit slice has 10 blocks with 5 transmission levels $T_6 - T_{10}$. Given a source of constant input intensity I_0 , the intensity incident on the spatial light modulator due to a typical slice within the unit cell is given by,

$$I_{wn} = I_0 T_0 T_n; \quad n = 6 \dots 10 \quad (6.1)$$

where n is the label denoting the block transmission level, and T_0 is the transmittance of the local input image intensity (assumed constant over the unit cell area). Notice that $I_{w6} > I_{w(n-1)}$.

When the MSLM is operated in the hardclipped thresholding mode, the switching characteristic discussed in Section 5.1 results, and, S and I_{th} (threshold intensity) are determined by I_{wn} (Equation 6.1). That is, for fixed input intensity I_0 such that $I_{w10} < I_{th}$, then all block intensities I_{wn} will be less than I_{th} and the cell slice will have uniform transmittance; the modulator readout light from the cell

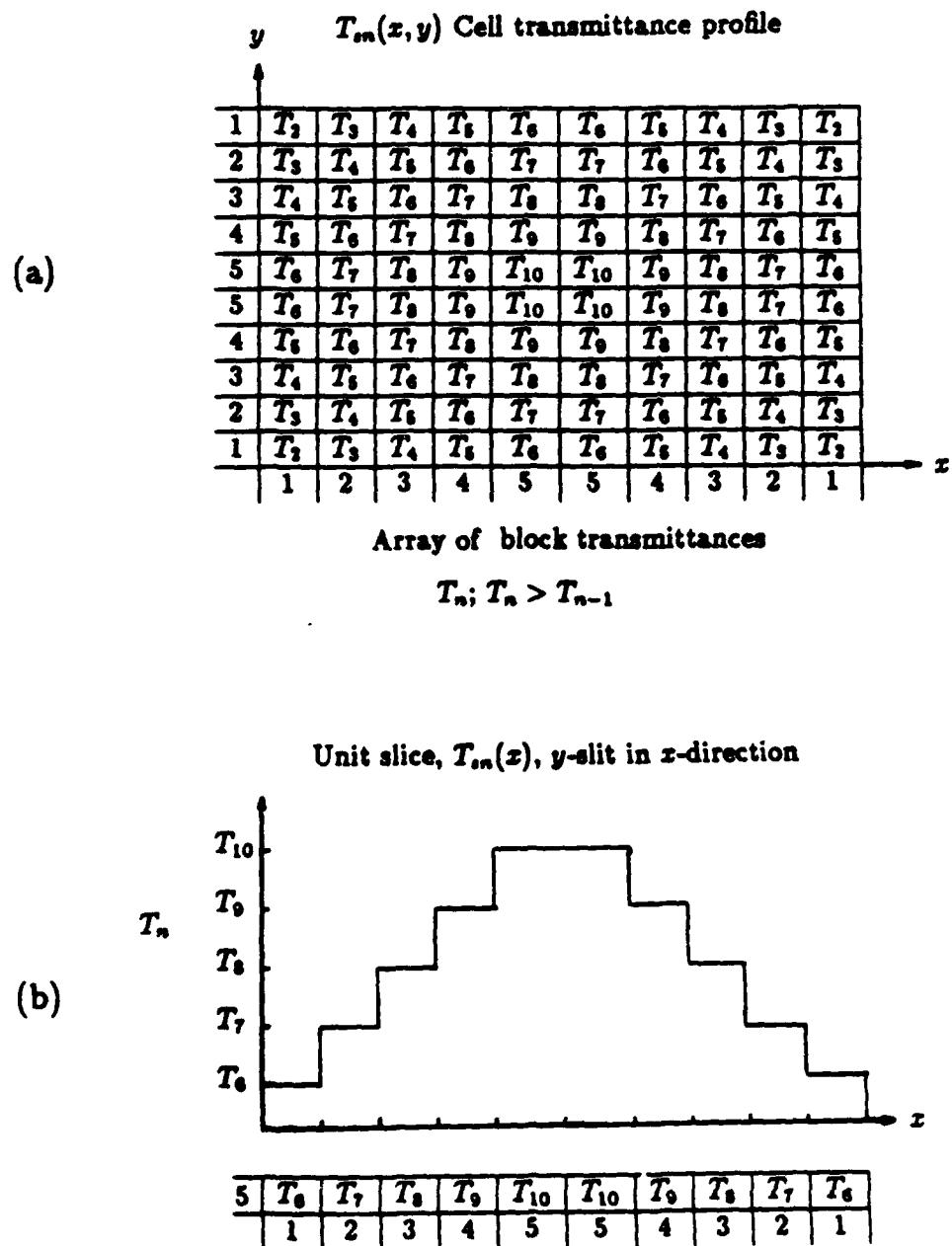


Figure 6.2: Profile of the contact screen utilized: a) 2-D profile b) y-slit extended in x-direction

I_{th} level	Grating carrier frequency f_{gn}, n block index
$I_{th} \geq I_{w10}$	$f_{g10} = 0$ unmodulated
$I_{w10} \geq I_{th} \geq I_{w9}$	$f_{g9} = \frac{1}{X}, \frac{1}{Y}$
$I_{w9} \geq I_{th} \geq I_{w8}$	$f_{g8} = \frac{1}{X}, \frac{1}{Y}$
$I_{w8} \geq I_{th} \geq I_{w7}$	$f_{g7} = \frac{1}{X}, \frac{1}{Y}$
$I_{w7} \geq I_{th} \geq I_{w6}$	$f_{g6} = \frac{1}{X}, \frac{1}{Y}$
$I_{w6} \geq I_{th}$	opaque

Table 6.1: Pulse frequency modulation: Grating carrier frequency corresponding to I_{th}

will be spatially unmodulated (grating carrier frequency f_{g10} of zero). Similarly, for $I_{w10} > I_{th} > I_{w9}$, the modulator readout light from that cell will be modulated by grating carrier frequency $f_{g9} = 1/X$. Further modulation results can be extended to 2-D and are summarised in Table 6.1.

When this pulse-frequency image is injected into a coherent optical processor, the diffraction spot at spatial frequency coordinate(s) $f_{x,y} = f\lambda f_{gn}$ in the Fourier plane will contain the information corresponding to points in the input image governed by f_{gn} .

Some example experiments that may be performed are as follows:

- A histogram of the original image is obtained by measuring the intensities of the spots in the Fourier plane.
- Perform equidensitometry by passing different orders in the Fourier plane.
- Enhance output image contrast by placing neutral density filters that have higher densities for lower intensities behind the diffraction spots and stopping all other diffraction spots.
- A pseudo-color image is obtained by using a white light readout source

instead of a monochromatic source and a set of color filters placed in the Fourier plane to selectively pass certain colors.

- Artificial stereo pictures are obtained by placing a set of optical wedges with different wedge angles behind the Fourier plane diffraction spots.

The following real-time experimental implementation of halftone screen processing was performed in collaboration with H. K. Liu of JPL. The experimental configuration is shown in Figure 6.3. The halftone images were written by imaging the input transparency/contact screen sandwich onto the MSLM by real-time hardclipped thresholding. Typical operating voltages quoted in Chapter 5 were used, and the threshold was adjusted by varying t_w and ramping V_s downward from 2.8 kV to 1.0 kV. The contact screen resolution was 65 lines/in., or equivalently, 26 lines/mm. This screen resolution was therefore around the resolution limit of the MSLM used. A $\times 5$ enlargement of the original screen was made by first making a color negative of the original, and then making a black and white negative by contact printing. This process preserved the relative gray scale profile of the contact screen.

The real-time halftone processing results are shown in Figure 6.4. Figures 6.4a and d show the thresholded images of an input gray scale transparency consisting of squared areas of increasing optical density for 2 different thresholds; while, b, c and e, f show the corresponding halftone images produced by sandwiching the input transparency first with the enlarged contact screen, and the original screen, respectively. Notice the spatial frequency modulation according to gray level; areas of higher incident intensity have lower grating frequencies than those of lower incident intensity. Also, areas where the intensity transmitted by the screen/transparency sandwich was lower than the threshold are unmodulated and

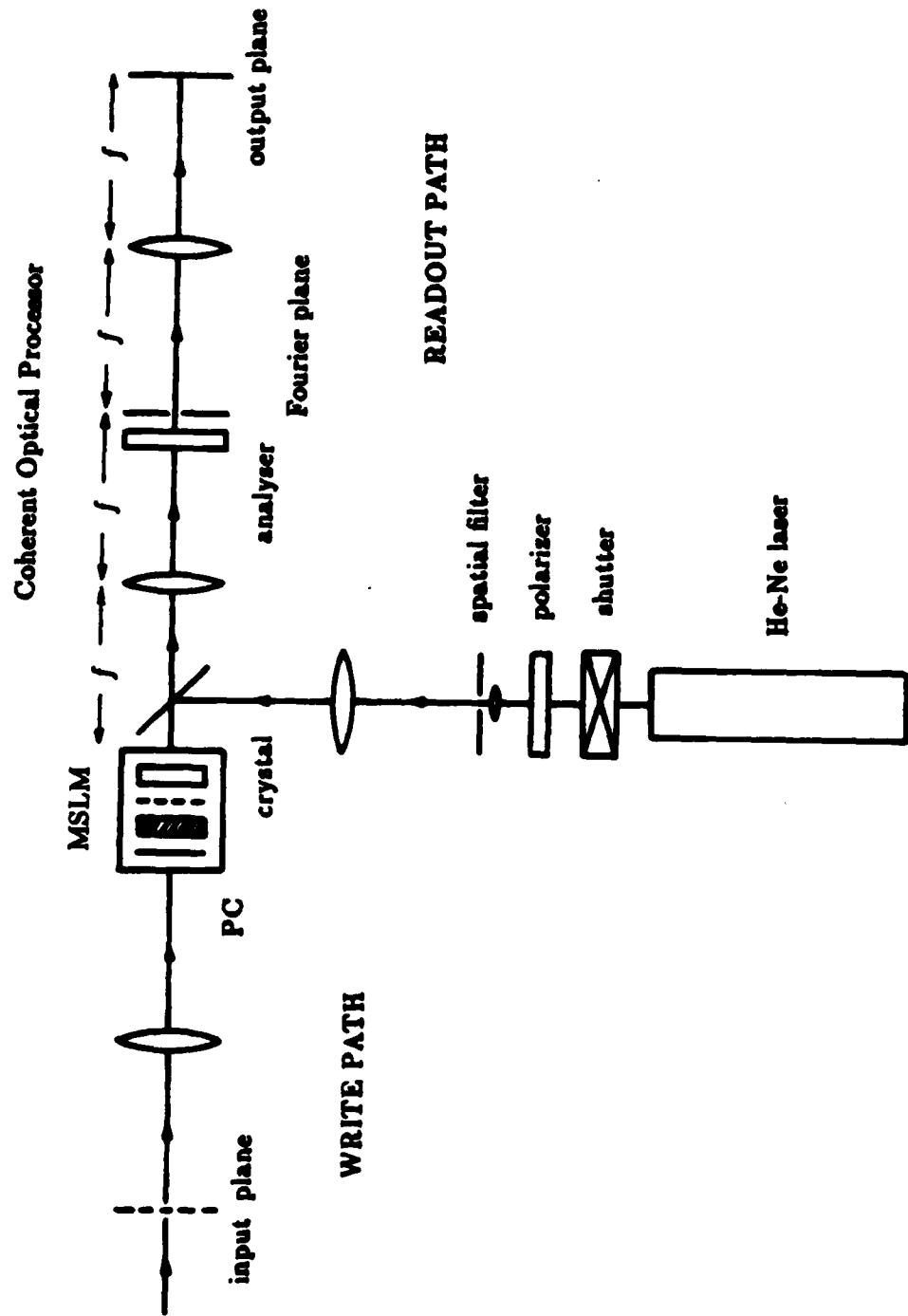


Figure 6.3: Optical System for implementing halftone processing

remain bright, while those where the intensity was higher than the threshold are dark. Figures 6.5a and b show the thresholded image of a second gray scale transparency of a woman's face and the corresponding halftone image. Write times used in these experiments ranged from 2.5 seconds to 1 minute.

Unfortunately the diffraction grating carrier frequencies which resulted from these halftone images were too closely spaced to be filtered successfully with the 30" focal length lenses that were available. Use of longer focal length lenses (1.5 - 2.0m) in the coherent optical processor will enable the mentioned non-linear operation to be realized.

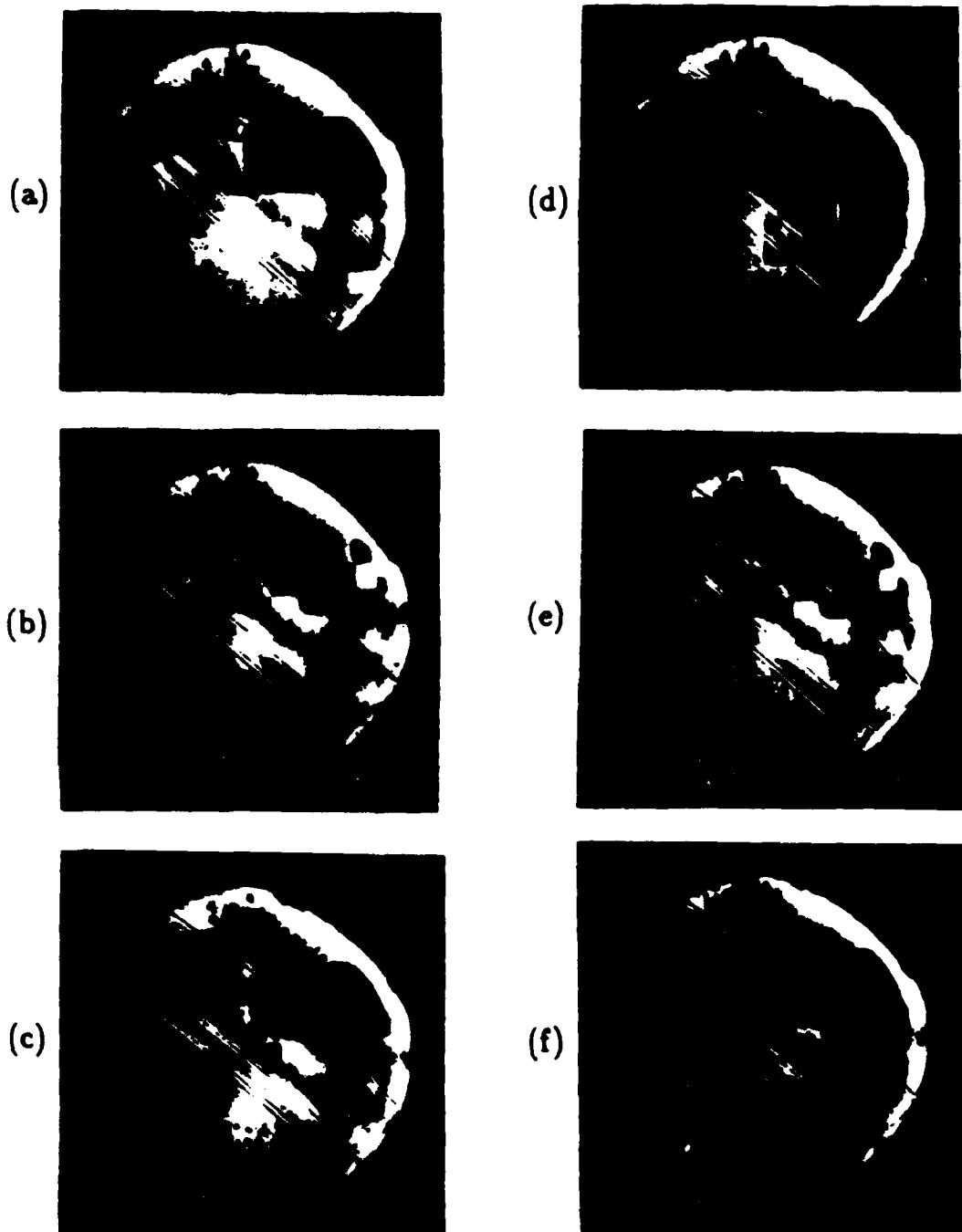


Figure 6.4: Real-time halftone screen processing results: a) & d) Thresholded gray scale image of squares for different threshold intensities, b), c) & e), f) Halftone images resulting from thresholding gray scale transparency/contact screen sandwiches for enlarged and original contact screens

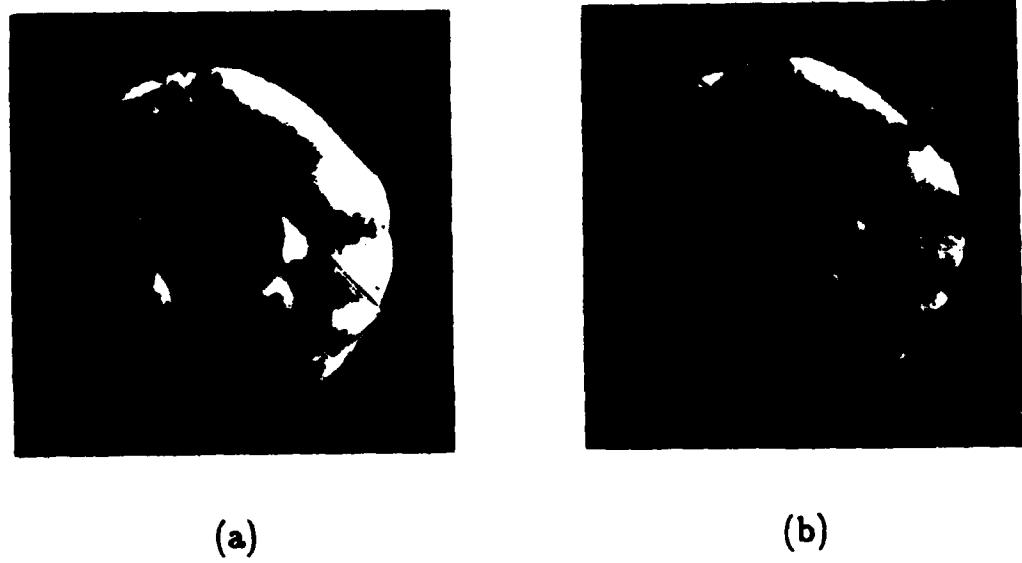


Figure 6.5: More real-time halftone screen processing results: a) Thresholded gray scale of a woman's face and b) corresponding halftone image

Chapter 7

Conclusion

7.1 Summary

The need for real-time, versatile spatial light modulators to replace photographic film has long been recognised. Such high-resolution, optically-addressed elements, capable of exploiting the inherent speed and parallel processing capability of optical processing, are essential for realizing futuristic optical computing architectures. This thesis described the techniques by which a standard MSLM can be modified and operated to achieve high-resolution, optically-addressed, variable-gamma characteristics applicable to a wide range of optical information processing applications.

The first such technique that was investigated, entailed converting the standard MSLM to a Fabry-Perot version MSLM. This required depositing dielectric mirrors with reflectivity R on both surfaces of the electro-optic crystal plate. This device was constructed and its high finesse output transmission characteristics (as determined by R) measured. The performance of the F-P MSLM was evaluated and its image processing limitations analysed. While the device built

had the desired high finesse behaviour, this high finesse capability imposed strict requirements on the crystal plane-parallelism necessary for a significant area of the crystal to modulate concurrently. These requirements were not met for the mirror reflectivity R that was used, and thus, imaging was not possible.

The next technique examined was the operation of the MSLM in a closed loop optical feedback configuration. The MSLM behaviour for both standard and F-P versions was modelled and analysed by numerical methods and was found to equilibrate under feedback. It is possible to compensate internal phase variations and to achieve optically bistable images using this configuration.

Next, the desired switching characteristic and the modes of operation that could be used to convert the periodic device characteristic were described and characterised. These modes included the electron/deposition grid stabilized mode and real-time hardclipped thresholding. The loss of proximity focusing and resulting image degradation in the electron deposition mode, make the hardclipped thresholding mode the preferred mode of operation. In the real-time hardclipped thresholding mode, the sensitivity of the MSLM was found to increase as a function of write intensity I_w and decrease as a function of write time t_w .

Finally, the applications of this high-resolution, optically-addressed, variable-gamma spatial light modulator in optical information processing were discussed in general. In particular, a standard MSLM operated in the hardclipped thresholding mode was used to demonstrate real-time halftone screen processing. Unfortunately, the spatial resolution of this device was not high enough to achieve spatial filtering in the transform plane which would have enabled further interesting processing operations.

The standard and the F-P MSLMs were found to have quite different charac-

teristics. The high finesse Fabry-Perot device is an intrinsically nonlinear, high-gamma modulator and can be applied to many nonlinear information processing applications in its linear operating modes. On the other hand, the standard MSLM may be applied to linear (low gamma) processing in its linear modes, and nonlinear (bistable) processing in its thresholding and closed-loop modes. This makes this particular modulator appealing since its versatility permits its use in many facets of a futuristic optical computer in addition to near-term applications for image and signal processing. Examples of present day implementations include radar, real-time pattern recognition, real-time holography and in industrial inspection systems.

Both the standard and the F-P MSLM can be operated in the optical feedback configuration to achieve internal phase compensation and optical bistability. This is advantageous since it allows the internal device phase deviations to be compensated and optical bistability may be achieved with a relatively simple architecture.

While the MSLM has now been studied and characterised at length, there still remain improvements to be made. The following section describes some suggestions for future work that have arisen during the course of completing this thesis.

7.2 Future Work

The research described in this thesis is the most recent work devoted to the development of the optically-addressed MSLM, here at M.I.T.. In the early years following its conception, most efforts were focused on investigating the fundamental operating characteristics and developing the optical architecture of the MSLM.

Later studies were more concerned with performance evaluation in terms of speed and resolution, and demonstration of optical information processing.

This thesis has attempted to study the operation of the MSLM in a methodical and complete manner. Modification of the standard MSLM to a F-P device was evaluated and several operating modes (closed and open loop) were characterised.

Future work on optically-addressed MSLMs should include demonstration of MSLM operation in the closed-loop feedback configuration. The models for this architecture have been calculated and discussed herein. Also, the MSLM should be used to implement halftone screen processing (as described in Chapter 6), and the resultant halftone images injected into a coherent optical processor with high quality, long focal length, Fourier transform lenses. Spatial filtering in the Fourier plane can then be employed to achieve many interesting image processing operations.

The Fabry-Perot MSLM needs to be developed further. Such a device with a visible photocathode is desirable in many nonlinear processing applications. The present polishing technology limits the achievable crystal face plane-parallelism, which in turn imposes a limitation on the maximum useful finesse of the F-P (related to mirror reflectivity R). This limitation on the etalon plane-parallelism may be alleviated by eliminating the use of a polished thin crystal plate and developing an MSLM which utilises thin film modulation materials instead. Such materials could be deposited uniformly onto a $\lambda/20$ optical flat by evaporation or some other technique, and would have substantially better plane-parallelism. These materials would also eliminate the need to be able to polish very thin flat crystals to achieve higher resolution devices, and might also reduce the cracking of the thin crystals due to strain and high electrostatic fields.

CHAPTER 7. CONCLUSION

101

Further studies of the limitations of presently available MSLMs are also necessary. These include investigation of the MTF, speed, usable finesse, maximum operating voltages and minimum thickness of the crystal plate.

Appendix A

Fabry-Perot MSLM Assembly

Considerable effort was devoted to designing a novel scheme for assembling a vacuum-demountable MSLM. The design which was implemented addressed the following problems encountered in previous designs:

1. the crystal crystal was epoxied and irretrievable
2. the glass substrate exerted pressure on the MCP through the grid
3. parallelism of components and even pressure were governed by the same teflon bolts which held the device in place
4. changing device components required disassembly of the entire device
5. the device was assembled directly onto the 6" vacuum flange for the vacuum cell.

The architecture which was developed for the construction of the Fabry-Perot MSLM was also utilised by Phillip Mak to build a electron-beam MSLM and is very compact in physical dimensions [19].

Component	Dimensions	inches	mm
MCP	Active area diameter	1.000	25.4
	Outer edge diameter	1.287	32.7
	Thickness	0.017	0.432
Crystal	Outer edge diameter	1.004	25.5
	Thickness	0.012	0.300
Grid	Inner ring diameter	0.700	17.8
	Outer ring diameter	0.876	22.3
	Ring thickness	0.055	1.31

Table A.1: Relevant dimensions of the Fabry-Perot MSLM components

Figure A.1 contains the individual flange specifications. The dimensions of the flanges were carefully chosen for use with the components of dimensions shown in Table A.1. Note the larger #6 nut clearance holes in the MCP and crystal flanges. Basically, the recessed lip in the center of the flanges allow the MCP, grid and crystal to be held in place without any epoxy, and the physical dimensions permit minimal pressure exertion on each component.

The metal flanges were manufactured using 0.0625" (1.59mm) thick, type 304 stainless steel. The MCP output electrode is made of 0.002" (0.051mm) nickel foil and the bolt holes machined to specifications between two 0.25" (6.35mm) slabs of aluminum to preserve its flatness. The insulating layers are made of 0.005" (0.127mm) teflon sheet.

This design allowed the device to be assembled in situ and then mounted onto the 6.0" (152mm) vacuum flange using the center *B*, #6 clearance holes, and teflon bolts (see Figure A.1). These bolts were only tightened as much as necessary to keep the device in place and exert no pressure on the components. The amount of time that the cell needed to be off-vacuum was thus minimized. The center grid flange served as an anchor for the MCP and the crystal/grid

APPENDIX A. FABRY-PEROT MSLM ASSEMBLY

104

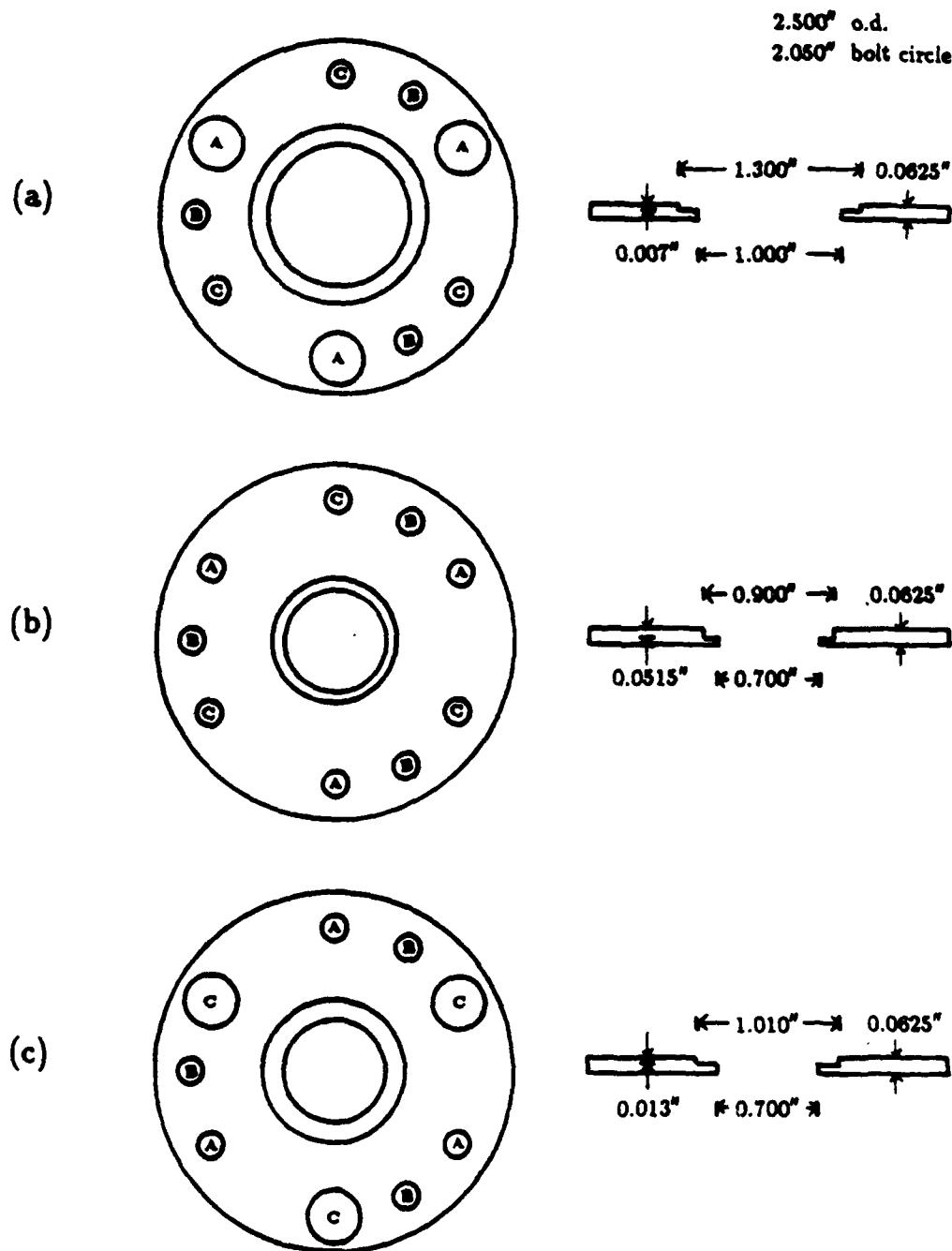


Figure A.1: Fabry-Perot flange specifications: a) MCP, b) Grid and c) Crystal flanges

sections independently; thus, the MCP could be changed without affecting the crystal or grid, and vice versa. This independence of components is enabled by the fact that the MCP mounting hardware utilised the C holes (larger clearance size hole in the crystal flange), while the crystal/grid hardware utilised the A holes (larger clearance size hole in the MCP flange). Corresponding holes are 120° apart on the bolt circle. None of the MSLM components were epoxied and could be removed and replaced. This feature is especially convenient for interchanging high and low strip current MCPs, as well as, Fabry-Perot and standard crystals. The MCP section can also be modified to accommodate a two MCP configuration by using the appropriate flanges with the correct bolt configuration.

The following procedure was used to build the Fabry-Perot device. All hardware and tools were degreased using reagent trichloroethane as necessary, and washed with reagent grade acetone and methyl. A final wash in semiconductor grade acetone completed the cleaning cycle.

Assembly steps are as follows:

1. *Crystal/grid assembly*

Determine crystal conducting surface

Place crystal in recess of crystal flange, electrode side down

Insert #6 - 32 nylon screws in the three A holes, posts up

Place one layer of teflon sheet on top of crystal flange

Center grid on teflon layer

Wrap grid wire around screw post

Position grid flange with grid in recess

Tighten bolts

2. MCP assembly

Insert #6 – 32 nylon screws in the three C holes, posts up

Wrap MCP output wire around screw post

Center Nickel electrode, attach wire

Place two layers of teflon sheet on Nickel electrode

Position MCP in center of teflon

Position MCP input flange with MCP in recess

Tighten bolts

3. Mount device on 6" flange

Mount device on vacuum flange using B holes

Overall, this design was quite successful. The independent nature of the MCP and the crystal/grid sections was actually tested when the the original grid had to be replaced. The only disadvantages were some movement of the crystals under high electric field and the active area of the crystal being reduced from 1.0" to 0.7" (that is, from 25mm to 18mm).

Appendix B

Anomalous Observations

While conducting the experimental work for this thesis, a number of unusual observations were made about the MSLMs. These anomalous characteristics have been categorised with reference to the particular device with which it was observed. A reasonable explanation is given wherever possible; however, in the case of the Hamamatsu device, much of the necessary information required to form a conclusive answer was unknown. Therefore, the explanations given were formulated partly from the limited information available and partly from some speculation on the author's part. These observations are included in the hope that they prove useful to other users of the MSLM.

B.1 Fabry-Perot MSLM

B.1.1 Marked Decrease in Device Speed

Slowing down of the device after the grid voltage was accidentally raised to +2.5 kV:

The high voltage resulted in a stretching of the center of the grid, so that, although there was no actual short to any other component, the transmission of

the grid was severely impaired by its flapping movement during operation, and resulted in less electrons reaching the crystal.

B.1.2 Bias Voltage Drifting

Rapid drifting of the bias voltage during operation (decreases about 600V in 45 minutes):

The high strip current MCP was heating up the crystal faster than heat could be dissipated (slow in vacuum), resulting in a rise in the net crystal temperature. Also, heat is generated from the stopping power of the crystal; that is, kinetic energy of the electrons is converted into heat $\sim eV$, where $V = |V_b + V_{grid} + V_g|$ (V_b is the MCP ejection energy $\sim 10V$, and V_{grid} and V_g are the grid and gap voltages, respectively). The electro-optic coefficients of the LiNbO₃ are temperature dependent and will change slightly, causing the bias voltage to shift.

B.2 Hamamatsu Standard MSLM

B.2.1 Unusually high $V_{\pi R}$

The reflex halfwave voltage $V_{\pi R}$ was measured around 4.5 kV:

Possible error in the crystal cut; large voltage drop across the dielectric mirror; inconclusive.

B.2.2 Nonuniform charge build-up

Charged tended to be more attracted to the edges of the crystal:

Possibly related to the internal device architecture; inconclusive.

B.2.3 Dark Ring

A dark ring was observed on the crystal output when read out (see Figure B.1):

The observed ring always remained dark and did not modulate when charge was added to the crystal surface. This ring has the appearance of an interference fringe and is perhaps the result of a reflection from the glass substrate interfering with the image.

B.2.4 Ghost Image

A ghost image of the resolution chart that had been previously stored in the device, was found to remain embedded in the dielectric mirror (Figure B.2) even after the crystal was erased:

This ghost image was found to result after a few hours of image storage, and did not disappear for a couple of weeks. Device usage seemed to accelerate the ghost image removal. The ghost modulated when the crystal voltage was changed with uniform incident intensity; and was superimposed in the background whenever any images were written onto the device.

This phenomenon probably stems in the materials used for fabricating the dielectric mirror. The storage time of the dielectric mirror seems too long and charge is trapped in its surface. This trapping phenomenon may also result from the presence of an additional secondary electron emission material evaporated on top of the dielectric mirror: charge may penetrate this material and become trapped by the dielectric mirror.

B.2.5 Image Curvature

A slight curvature of the image recorded on the MSLM was observed in the output

APPENDIX B. ANOMALOUS OBSERVATIONS

110

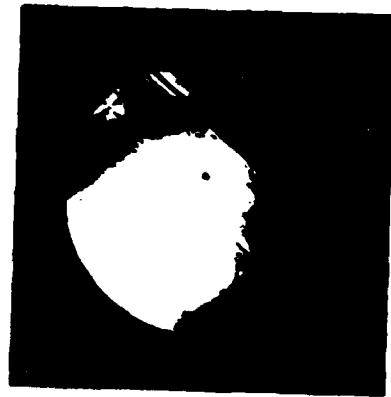


Figure B.1: Dark ring



Figure B.2: Ghost image

plane:

Figure B.3a is a photograph of the focused input image of the Air Force resolution chart that was incident on the photocathode; notice that there was no distortion. Figure B.3b is a photograph of the image produced when the resolution chart was placed against the output window of the MSLM (the interference fringes are the result of reflections from the front and back faces of the quartz plate); notice this image is also undistorted. Since the curvature was not introduced by either the write or the readout path, it must have been introduced by the device (see Figure B.3c). Upon closer inspection, the photocathode was found to be slightly convex, which would account for the image curvature. This effect can be corrected by using a fiberoptic faceplate in front of the photocathode. The image is then focused on the input planar face and transferred to the photocathode.

B.2.6 Crystal Cracking

Cracking of the crystal along the z-axis:

This was perhaps the most puzzling observation of all. It was found that there were two distinct cracking phenomenon; one seemed to be reversible, while the other was permanent.

Figure B.4a is a photograph of the output image of the Hamamatsu MSLM in the condition in which it was received at M.I.T. The conditions under which this initial crack was formed are unknown.

However, during subsequent use, the following cracking of the crystal was observed:

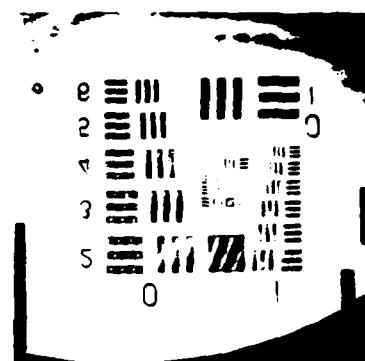
- *Positive Charge Deposition*

APPENDIX B. ANOMALOUS OBSERVATIONS

112



(a)



(b)



(c)

Figure B.3: Image curvature: Resolution chart a) imaged by write optics, b) imaged by readout optics, and c) stored and read out of device

The following operations were performed with uniform incident intensity in all cases.

1. V_t was slowly decreased from 0 to -1875 V (2875 V of positive charge on the mirror); and then slowly returned to 0 V . The resulting output is shown in Figure B.4b, these observed lines are all parallel to the crystal axis.
2. The crystal voltage was increased slowly from 0 to $+2840\text{ V}$ and then reduced again to 0 V . The severity of the observed lines were found to have decreased (see Figure B.4c).
3. The device was shut down and the crystal photographed 2.5 hours later (see Figure B.4d).

Notice that the *lines* seem to slowly disappear with time. This recovery is accelerated by depositing and removing negative charge. The above phenomenon was repeatable. If the crystal voltage was lowered to only -1200 V , the lines also appeared, but were not as severe.

• Negative Charge Deposition

Again, the following occurred with uniform incident intensity in all cases.

1. V_t was accidentally raised to about $+3300\text{ V}$ (2300 V of negative charge), then lowered to 0 V . Notice, the single crack which appeared across the crystal, parallel to the x -axis and the initial crack (see Figure B.4e). The faint lines in the background resulted from positive charge deposition and were in the process of fading away. This new crack was found to modulate more slowly and out of phase with the remainder of the crystal. Similar behaviour was also observed for the original crack.



(a)



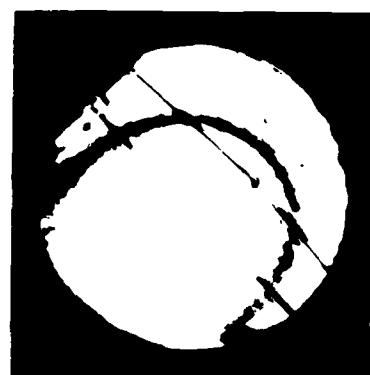
(b)



(c)



(d)



(e)

Figure B.4: Cracking phenomenon: a) Original crack
Positive charge deposition: b) V_b , 0 to -1875 to 0 volts, c) 0 to $+2840$ to 0 volts
and d) after 2.5 hours
Negative charge deposition: e) V_b accidentally raised to $+3300\text{ V}$

This crack did not fade or recover as in the previous case, but instead, remained on the crystal.

2. The voltage was raised to +3050 V a few months later, and yet a third line appeared with the same characteristics.

The above cracks were permanent and severe. They were parallel to the crystal axis and modulated more slowly than the background. These lines could also be observed on the crystal surface by looking directly into the output window of the device.

The possible sources of this phenomenon are the dielectric mirror or the crystal. However, the parallelism of these cracks to the crystal axis indicates that the explanation lies with the crystal. In addition, the dielectric mirror is deposited without any uniform orientation and consists of alternating layers of two different materials; thus, eliminating any reason for the uniformly formed lines.

Since these effects have not been observed in vacuum demountable devices with 300 - 330 μm crystals, the cracking is probably related to the fact that the crystal thickness has been reduced to 50 μm for these sealed-off devices.

Without further, possibly destructive experimentation, the following explanation seems to be the most feasible. Consider the 50 μm thick 55°-cut LiNbO₃ crystal, epoxied to a glass substrate in Figure B.5a. If positive charge is deposited on the dielectric mirror the resulting voltage V_z across the crystal produces an electric field across the crystal (from right to left). This field tries to rotate the z-axis of the crystal, so that it is parallel to the field; this would result in poling of the crystal (Figure B.5b). Conversely, if negative charge is deposited, the resulting electric field tends to rotate the z-axis in the opposite direction (Figure B.5c).

Sometime in its manufacture, the LiNbO₃ crystal was grown and its individual

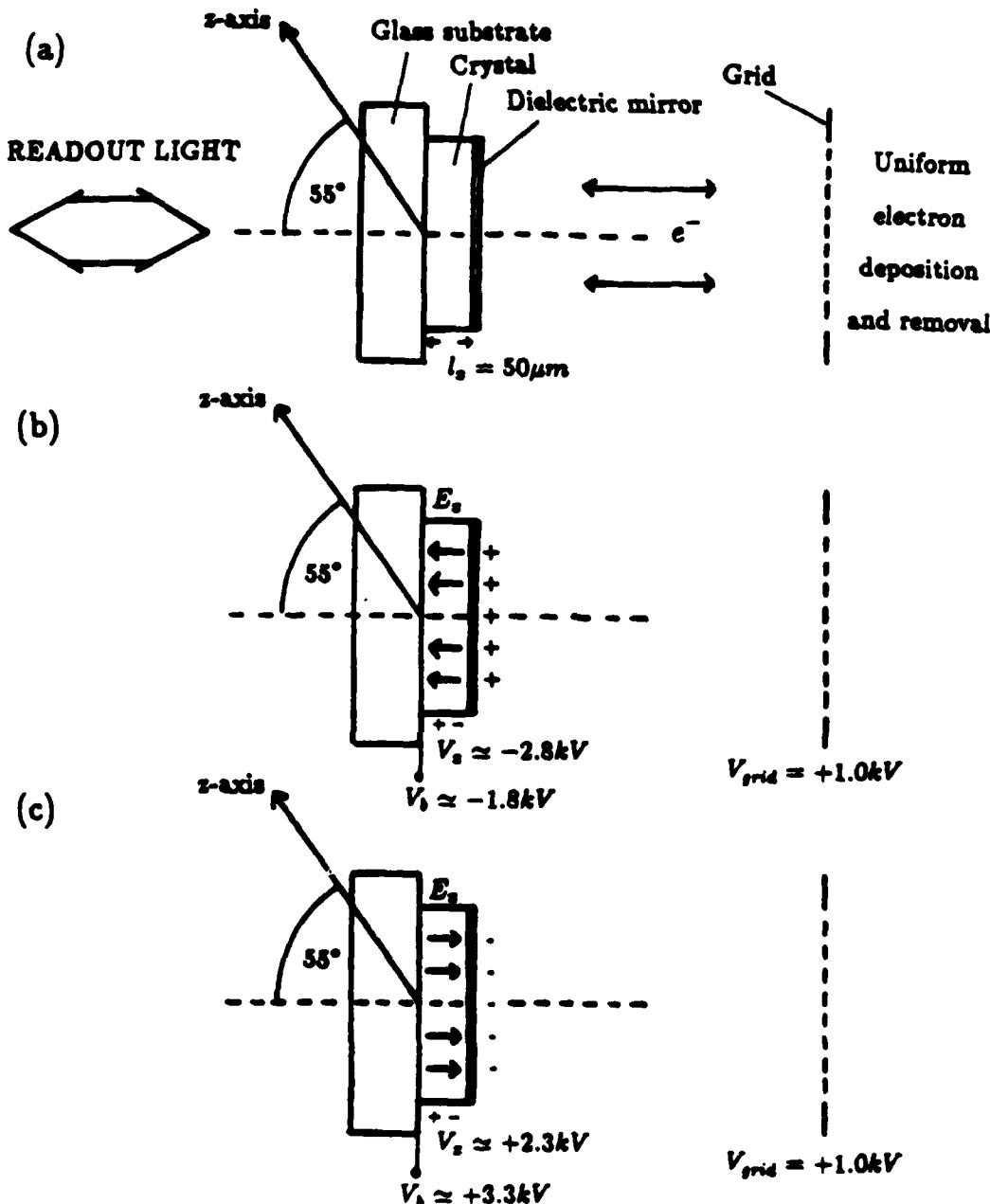


Figure B.5: Charge-induced poling explanation of oblique-cut LiNbO_3 : a) Crystal orientation in MSLM, b) positive charge deposition (electron removal), and c) negative charge deposition (electron deposition)

domains rotated so that their z -axes were all parallel. The electric fields that are created by deposited positive and negative charge attempt to rotate the z -axes of these individual domains; that is, pole the crystal in the transverse direction. The forces necessary for a *thick* crystal are much greater than those for a $50\text{ }\mu\text{m}$ crystal.

Notice that with positive charge deposition, the tendency is for the charge-induced electric field to realign the z -axis by rotating it 55° and thus, is an easier operation than when negative charge is deposited and the z -axis wants to rotate 125° . It seems reasonable that the positive charge case is somewhat reversible, since once the charge is removed the domain will tend to realign itself. However, with negative charge deposition, the rotation of a domain may result in cracks and permanent physical damage in the crystal and is thus irreversible; but since a larger angle of rotation is required, fewer domains are affected. Domains which do rotate have z -axes lined up with the induced field and have the halfwave voltage corresponding to z -cut LiNbO_3 (about 3100 V). Thus the modulation would be slower than the background 55° -cut crystal ($V_{zR} = 1250\text{ V}$).

While this explanation is supported by the observations described above, further studies of the effects of strong electric fields on thin polished electro-optic crystals need to be conducted.

Bibliography

- [1] C. Warde, A. M. Weiss, A. D. Fisher and J. I. Thackara, "Optical information processing characteristics of the microchannel spatial light modulator," *Applied Optics*, vol. 20, no. 12, pp 2066-2074, 1981.
- [2] C. Warde and J. I. Thackara, "Oblique-cut LiNbO₃ microchannel spatial light modulator," *Optics Letters*, Vol. 7, pp 344-346, 1982.
- [3] A. M. Weiss, "Development of the Microchannel Spatial Light Modulator," *SM Thesis*, Massachusetts Institute of Technology, December 1980.
- [4] J. I. Thackara, "Image Processing Characteristics of the Microchannel Spatial Light Modulator," *SM Thesis*, Massachusetts Institute of Technology, June 1982.
- [5] A. D. Fisher, "Techniques and Devices for High-Resolution Adaptive Optics," *PhD Thesis*, Massachusetts Institute of Technology, August 1981.
- [6] J. A. Kottas, "Optical Pattern Processors for Matrix Multiplication and Symbolic Inference," *SM Thesis*, Massachusetts Institute of Technology, February 1986.
- [7] A. Schwartz, X. Y. Wang and C. Warde, "Electron-beam-addressed microchannel spatial light modulator," *Optical Engineering*, Vol. 24, no. 1, pp

119-123, 1985.

- [8] J. L. Wiza, "Microchannel Plate Detectors," *Nuclear Instruments and Methods*, Vol. 162, pp 587-601, 1979.
- [9] T. Hara, Y. Ooi, T. Kato and Y. Suzuki, "Microchannel Spatial Light Modulator with improved resolution and contrast ratio," in *Proc. SPIE Nonlinear Optics and Applications*, Vol. 613, pp 153-157, 1986.
- [10] C. Warde, H. Lamela-Rivera and H. K. Liu, "High-Gamma spatial light modulator for nonlinear optical processing," in *Proc. SPIE Nonlinear Optics and Applications*, Vol. 613, pp 153-157, 1986.
- [11] J. W. Goodman: *Introduction to Fourier Optics*, McGraw-Hill, 1968.
- [12] S. H. Lee: *Optical Information Processing*, Springer, 1981.
- [13] F. T. S. Yu: *Optical Information Processing* Wiley-Interscience, New York, 1983.
- [14] H. Kato and J. W. Goodman, "Nonlinear filtering in coherent optical systems through halftone screen processes," *Applied Optics*, Vol. 14, no. 8, pp 1813-1824, 1975.
- [15] T. C. Strand, "Non-monotonic, non-linear image processing using halftone techniques," *Optics Communications*, Vol. 15, no. 1, pp 60-65, 1975.
- [16] H. K. Liu, J. W. Goodman and J. L-H. Chan, "Equidensitometry by coherent optical filtering," *Applied Optics*, Vol. 15, no. 10, pp 2394-2399, 1976.
- [17] H. K. Liu, "Coherent optical analog-to-digital conversion using a single halftone photograph," *Applied Optics*, Vol. 17, no. 14, pp 2181-2185, 1978.

BIBLIOGRAPHY

120

- [18] D. Zhao, C-K Chiang and H. K. Liu, "Contact screen image subtraction technique," *Applied Optics*, Vol. 20, no. 24, pp 4234-4238, 1981.
- [19] P. Mak, "Compact design of an Electron-beam-addressed Microchannel Spatial Light Modulator," *SB Thesis*, Massachusetts Institute of Technology, February 1987.

APPENDIX C

**A Comparison of the Hughes and the Loral
LCLV Characteristics**

by

Jeffrey A. Davis

and

Roger A. Lilly

**San Diego State University
San Diego, CA 921820**

TABLE OF CONTENTS

Section	Page
I. Introduction	3
II. Operating Principles of Liquid Crystal Light Valves	5
III. Performance of the Hughes LCLV	17
IV. Performance of the Loral LCLV	40
V. Discussion	67
VI. References	70

I. INTRODUCTION

There has been a great deal of activity in the development of compact real-time coherent optical correlators for use in a variety of pattern recognition applications (1-3). The most critical component in many correlator designs is a liquid crystal light valve (LCLV) which allows an incoherent input image from an optical viewing system to be converted into a coherent image for subsequent analysis by the optical correlator.

The operation of these devices depends on the electrical control of the birefringence of a thin liquid crystal layer. The devices which we have studied operate in a reflection mode in which the linearly polarized readout light enters from one side, passes through the liquid crystal, reflects off a mirror, and returns. The birefringence induces a component of polarization perpendicular to the incident readout polarization. This perpendicular polarization is the output signal of the LCLV. The degree of birefringence is controlled by having a writing beam incident on a photoconductor layer. As the intensity of the writing light increases, the resistance of the photoconductive layer decreases, causing a greater fraction of the applied voltage to appear across the liquid crystal layer.

Currently, there are two manufacturers of these liquid crystal light valves; Hughes Aircraft and Loral EOS. The Hughes device was introduced in the 1970's (4,5) and has been in commercial production for a number of years. Our tests were

performed on a unit with serial number 11185 obtained from Hughes as a donation to our educational program. Although the Loral device was also introduced in the 1970's, it is still in development stages. The one we are using has the serial number 5B and is on loan from Loral EOS. This report summarizes those results.

The next section gives a general discussion of the operation of these devices and analyzes their common points as well as their differences. Then, the next two sections discuss the experimental results for the two devices. Finally, we summarize our findings and compare the characteristics of the two devices.

II. OPERATING PRINCIPLES OF LIQUID CRYSTAL LIGHT VALVES

A drawing showing the construction of the Hughes liquid crystal light valve is shown (4) in Figure 1. The Loral device, although very similar, has several important differences which will be discussed later.

The Hughes device consists of 4 major components: a Cadmium Sulfide (CdS) photoconductor, a Cadmium Telluride (CdTe) light blocking layer, a dielectric mirror, and a nematic liquid crystal with an AC bias voltage placed across the entire package.

Normally, the majority of the voltage appears across the photoconductive layer when the device is in the off state (no incident writing light). In practice, there is some leakage and a small fraction of the voltage does appear across the liquid crystal layer. Note that the AC bias voltage produces an average DC bias across the liquid crystal since the photoconductor layer and the light blocking layer combine to form a diode junction. The use of AC also enhances the device lifetime.

When light is incident on the photoconductor, the effective resistance of the photoconductor decreases causing the fraction of the bias voltage appearing across it to decrease. This results in a greater fraction of the bias voltage appearing across the liquid crystal layer. This light beam which is incident on the photoconductive layer is referred to as the WRITING beam. Obviously the wavelength of the writing beam should be chosen to affect the maximum change in the

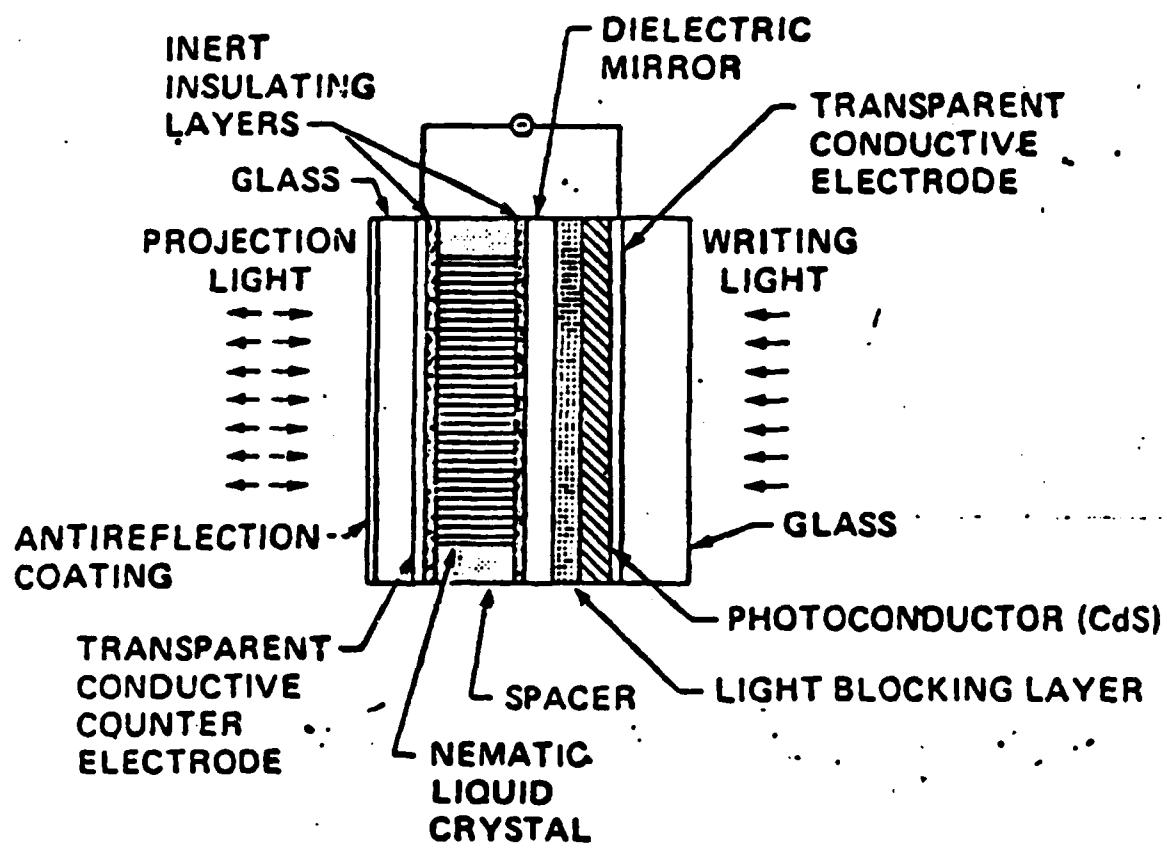


Figure 1 - Schematic drawing of the Hughes liquid crystal light valve. (From Ref. 4)

photoconductive layer for a minimum writing intensity and depends on the particular photoconductive material used. The photoconductive layer in the Hughes device is CdS while As₂Se₃ is used in the Loral device. Therefore we would expect different wavelength sensitivities for the two devices.

The liquid crystal is a twisted nematic type whose molecules have a long cigar shape. In the absence of an applied electric field, the axis of symmetry of the molecules is oriented parallel to the entrance face of the liquid crystal. The symmetry axes of the molecules undergo a relative twist when followed through the liquid crystal layer as shown (7) in Figure 2a. The orientation of the molecules at either end is determined by "twining" the substrate boundaries. This term means that grooves are milled into the surface layers which force the molecules to orient in a desired direction at the two surfaces. In the Hughes device, the molecules twist through a 45 degree angle while the Loral device uses a 90 degree twist.

As the voltage across the liquid crystal layer increases, both the twist and tilt of the liquid crystal molecules change as shown in Figure 2b and 2c which affects the birefringence of the liquid crystal layer. Figure 3 shows graphical plots for both the twist and tilt as the electric field changes (4).

With no voltage applied to the device, the twist increases uniformly across the liquid crystal thickness. However, as the voltage increases, the liquid crystal molecules untwist and become aligned with the preferred directions at the respective

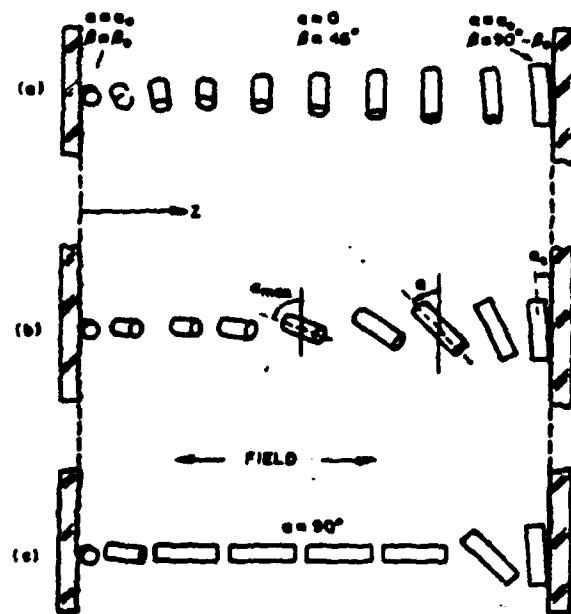


Figure 2 - Drawing showing the orientation of the liquid crystal molecules in a 90 degree twist device. Figure 2a shows the molecular orientation with no applied field. Figure 2b shows the effects of a moderate field while Figure 2c shows the effects of a large electric field. (from Ref 7).

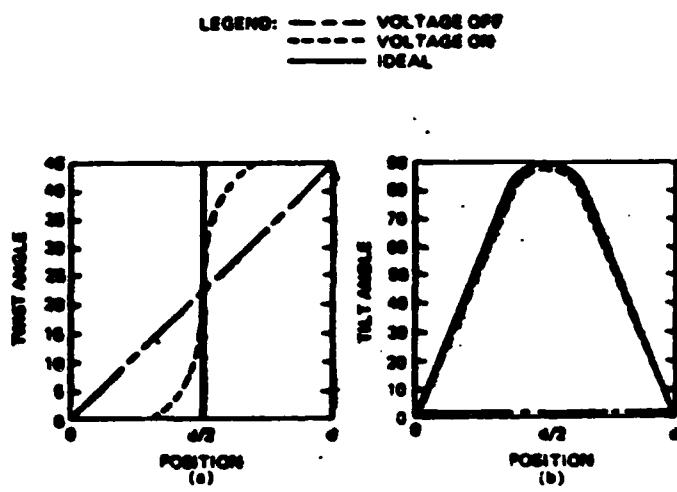


Figure 3 - Calculated values showing molecular orientation as a function of position across the liquid crystal cell for a 45 degree twist device. Figure 3a shows the twist while Figure 3b shows the tilt. (From Ref. 4).

edges as shown in Figure 3

The tilt is initially zero and the long cigar shaped molecules are aligned as shown in Figure 2a. However, when the electric field is increased, the molecules gradually tilt as shown in Figures 2b and 2c with the long axis becoming parallel to the applied field and hence parallel to the direction of propagation of the light incident on the liquid crystal. This incident light is called the READING beam and is initially polarized parallel to the preferred orientation direction of the liquid crystal molecules on the input side.

Note that the voltage on the liquid crystal layer can be increased by either increasing the total bias voltage across the device or by having the writing beam decrease the resistance of the photoconductive layer.

If there is no bias voltage across the liquid crystal layer, the plane of polarization of the reading beam rotates by following the relative twist of the major axes of the molecules. The light then reflects off the dielectric blocking layer and the polarization untwists as it returns. Therefore, the polarization of the reflected beam is exactly the same as the incident beam.

However, when an electric field is placed on the liquid crystal resulting in the configuration shown in Figure 2b, the polarization state of the incident reading beam remains unchanged through the first portion of the crystal since the twist has been removed. The symmetry axes of the liquid crystal molecules in the second half of the liquid crystal layer are now aligned at 45

degrees to the initial polarization direction (in the case of the Hughes device). Therefore, the second half layer becomes birefringent and produces an orthogonal polarization state as the beam passes through the liquid crystal layer, reflects, and returns. This polarization state can be separated by using a polarization splitting beamsplitter.

The relative birefringence depends on the propagation direction of the incident beam. If it is not orthogonal to the plane of the device, the effective thickness of the liquid crystal layer will depend on the incident angle with different birefringent effects for different angles. For this reason, normal incidence is recommended.

As the voltage increases further resulting in the configuration shown in Figure 2c, the major axes of the liquid crystal molecules become aligned to the propagation direction. Thus the liquid crystal layer is no longer birefringent and the polarization state of the reflected light is unchanged.

Therefore, the percentage of the orthogonal polarization developed by the device is initially zero, increases to some maximum, and then decreases again as the voltage across the liquid crystal layer increases as shown in Figure 4 (4).

The operation of a device having a 90 degree twist is similar with an important difference. Under the applied electric field, the twist of the molecules again aligns to the preferential directions at either end as shown in Figures 2 and 3. However in this case, the relative twist in the ON state is

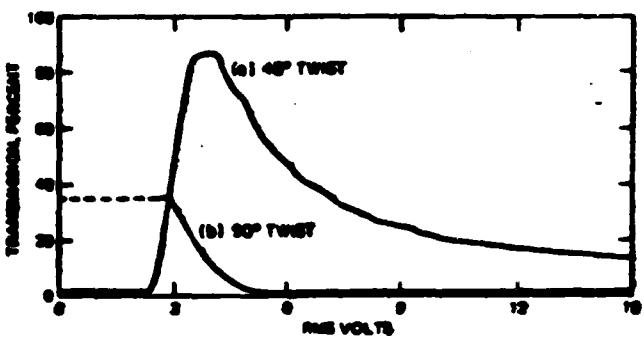


Figure 4 - Experimental curves showing the relative transmission of liquid crystal devices with a 45 degree twist and a 90 degree twist. (From Ref. 4)

90 degrees. Therefore, the polarization state of the reading beam remains unchanged through the first half of the liquid crystal layer since the twist has been removed as before. However, the symmetry axes of the molecules in the second half of the liquid crystal layer are now orthogonal to those in the first half. Therefore, no polarization state would be developed perpendicular to the initial polarization direction since the initial state of polarization would lie along the minor index ellipsoid axis of the liquid crystal layers in the second half of the device. The percentage of the orthogonal polarization state which is developed should therefore be less, and would only result from the variation in twist at the center of the liquid crystal layer as shown in Figures 2 and 3. Figure 4 (taken from Reference 4) shows the relative transmission of liquid crystal devices having 45 and 90 degree twists demonstrating the superiority of the 45 degree design. However, our experimental results show the reflectivity of the Loral device to be roughly comparable to that of the Hughes device in spite of the fact that the Loral device is reported to have a 90 degree twist. We do not understand this discrepancy.

As mentioned above, a greater fraction of the reading beam will be reflected into the opposite polarization state as the birefringence of the liquid crystal layer increases. Obviously in the optimum case, 100% of the reading beam would be converted into the orthogonal polarization state. In our experiments, we measured the absolute device reflectivity.

An important difference in operating parameters for the two devices is in the AC frequency of the bias voltage which must be chosen to maximize the impedance match between the liquid crystal and the photoconductor/photocapacitor substrate (8). The recommended bias frequency for the Hughes device is 10 kHz. However, Loral specifies bias frequencies in the 10 - 50 Hz range.

This extremely low value for the Loral operating frequency raised some concerns about the effects of this AC bias on the response time of the device. The response time of the device depends on the response times of both the photoconductor and liquid crystal layers. Usually, the latter dominates and is dependent on the reorientation time required for the liquid crystal molecules under the applied electric field.

While the changes in the bias voltage at the operating frequency of the Hughes device are much faster than these reorientation times, this is not the case with the Loral device.

In the case of the Loral device, we were initially concerned that the response time of the molecules to the changing low frequency bias voltage might mask changes in the output due to changes in the input light intensity. As we will see, these concerns were justified.

There is another substantial difference between the two devices. The Loral device does not have an optical blocking layer placed between the photoconductor and liquid crystal layers while the Hughes device does. The purpose of this layer is to prevent the reading light beam from passing through to the

photoconductor and essentially becoming a writing beam. In the case of the Loral device, this omission appears to be motivated by economical considerations. The price of a device made without a blocking layer will be lower. However, the reading beam must be filtered since wavelengths longer than 600 nm will be transmitted through the dielectric mirror and will go through to the photoconductor. This can present substantial disadvantages since laser diode and convenient gas lasers operate at red or infrared wavelengths.

The optimum wavelength for the reading beam will depend on the type of liquid crystal used as well as the thickness of the liquid crystal layer. For optimal operation, the device should act essentially as a half wave plate converting the initial linearly polarized light completely to the orthogonal polarization. The total birefringence of the device can be varied by increasing the thickness of the liquid crystal layer. This thickness can be altered for optimum response at a desired wavelength. However, the thickness cannot be increased indefinitely since the time response of the device becomes slower with increasing thickness.

Patterns can be written onto the liquid crystal since the transverse photoconductivity is low enough to allow a spatial pattern to be imposed. Therefore, the locations of the birefringence will correspond with those locations where the photoconductor has been affected by the writing beam. However, this introduces spatial resolution or the modulation transfer

function (MTF) as another performance parameter.

As the period of a spatial pattern decreases, the separation of the low resistance portions of the photoconductor layer decreases along with the separation of the portions of the liquid crystal layer which are subjected to the higher bias voltage. However, there is some deterioration of the transverse photoconductivity, which limits the ultimate spatial period obtainable. This will be shown by a decrease in the resolvable spatial period on the reading side as the writing spatial period is decreased.

A variety of tests measuring these parameters were performed on the two devices including the effects of changing the bias voltage, bias frequency, wavelength of the writing beam, risetime and falltime, MTF, and wavelength of the reading beam. The results of these tests will be discussed for each device.

III. PERFORMANCE OF THE HUGHES LCLV

A. Introduction

This chapter documents the experimental measurements made on the performance characteristics of the Hughes liquid crystal light valve. Where appropriate, experimental details will be presented. Since little operating information was given on the Loral device, we decided to examine the Hughes device first to see the motivation for their operating specifications. Then, we would be able to optimize the operating specifications for the Loral device.

B. General Experimental Set-up

The general experimental set-up is the same as reported earlier (4). An Argon ion laser with an output at 514.5 nm was used as the writing beam since previous work indicated that the CdS photoconductive layer is sensitive to this wavelength. The laser beam was spatially filtered, expanded, collimated, and apertured to achieve a uniform intensity distribution across a portion of the photoconductor layer.

The reading beam was obtained from a Helium Neon laser at a wavelength of 632.8 nm which was also spatially filtered, collimated and apertured. In this case, the aperture diameter was smaller than the writing beam diameter to insure that the cross sectional area of the reading beam was smaller than the cross sectional area of the writing beam. In addition, care was taken to make sure the two beams spatially overlapped.

The linearly polarized Helium Neon laser beam passed through a polarization beamsplitter cube. The orthogonal polarization state of the reflected light was then separated by this beamsplitter and directed to a photodetector. The absolute incident light intensity on the LCLV was measured by placing the detector directly in front of the LCLV or by inserting a quarter wave plate between the beamsplitter and the LCLV.

It is critical that the initial polarization direction of the reading light be parallel to the direction of the liquid crystal molecules. This was determined by rotating the LCLV until the reflected light was nulled.

C. Bias Voltage

We first examined the operating characteristics of the device as a function of bias voltage. With the writing beam turned off, the absolute reflectivity of the Hughes device at the Helium Neon laser wavelength of 632.8 nm was measured as a function of bias voltage. Experimental results are shown in Figure 5. The frequency of the AC voltage was chosen to be 10 kHz in accordance with recommendations from the manufacturer. However, we found little variation in the performance of the light valve as this frequency was changed.

Several characteristics are seen in Figure 5. The reflectivity decreases to a minimum, goes back through a small peak, decreases again, and then goes through a large peak as the bias voltage increases. In order to minimize the reflectivity when there is no writing beam (as is usually desired), a bias of 10.3 VRMS is used. However, if a maximum response is desired in the absence of any writing beam, a bias voltage of 3.76 VRMS would be used. The consequences of this will be noted later.

The reflectivity is again plotted versus the bias voltage in Figure 6 for a writing beam intensity of 58 microwatts/cm² at the green 514 nm wavelength of the Argon ion laser. Although the curve is similar, the peaks have moved to lower voltages. In this case, the fraction of the bias voltage which appears across the liquid crystal layer increases due to the presence of the writing beam.

Therefore if the bias voltage is chosen to be 10.3 VRMS, the

reflectivity of the device will increase as the writing beam intensity increases. This will be examined next.

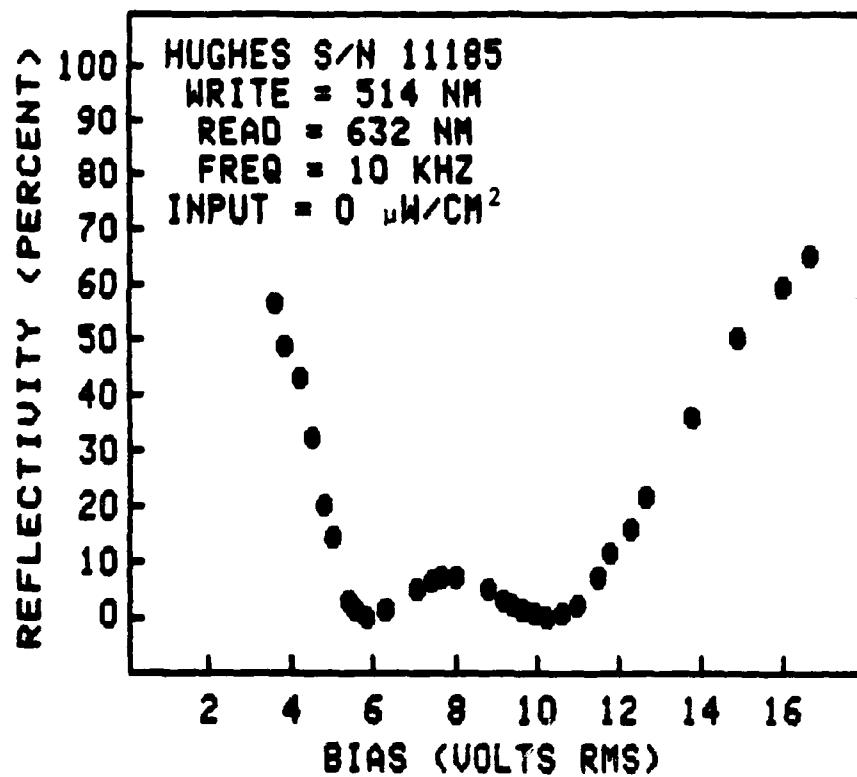


Figure 5 - Reflectivity of the Hughes LCLV at a wavelength of 632.8 nm as a function of AC bias voltage applied across the device. Here, the writing beam is turned off.

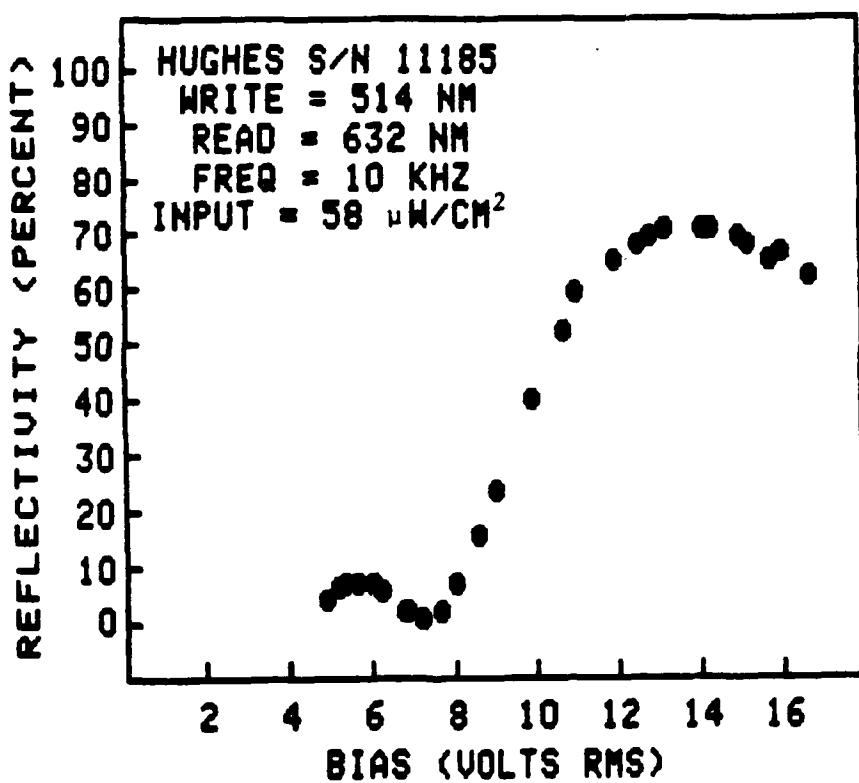


Figure 6 - Reflectivity of the Hughes LCLV at a wavelength of 632.8 nm as a function of AC bias voltage applied across the device. Here, a writing beam of 58 microwatts/cm² at the green wavelength of 514.5 nm is used.

D. Writing Sensitivity

Measurements were made of the absolute reflectivity of the Hughes device at a reading wavelength of 632.8 nm as a function of the intensity of the writing beam at 514.5 nm. Figure 7 shows the absolute device reflectivity as a function of writing intensity using an AC bias voltage of 10.3 VRMS and shows an increase in reflectivity with increasing writing intensity. In this experiment, the intensity incident on the LCLV was first measured. Then the absolute intensity reflected from the LCLV and passed by the beamsplitter was measured. This was taken as the absolute reflectivity, assuming a perfect beamsplitter. Note that moderately low intensities of 50 - 100 microwatts/cm² are able to achieve high reflectivities of 60 - 80%.

An interesting effect can be obtained by changing the bias voltage to 3.76 VRMS. Here, maximum reflectivity is obtained for zero writing input. Increasing the writing intensity again causes the curve in Figure 5 to move to the left. However, for this choice of bias voltage, the reflectivity will now decrease with increasing writing intensity as shown in Figure 8.

When an intermediate bias voltage is chosen, the reflectivity will vary as the curve in Figure 5 again to the LEFT relative to that bias point. Figure 9 shows the results for a bias voltage of 6.7 VRMS. In this case, the maximum reflectivity is only 7% and peaks at an intensity of 30 microwatts/cm².

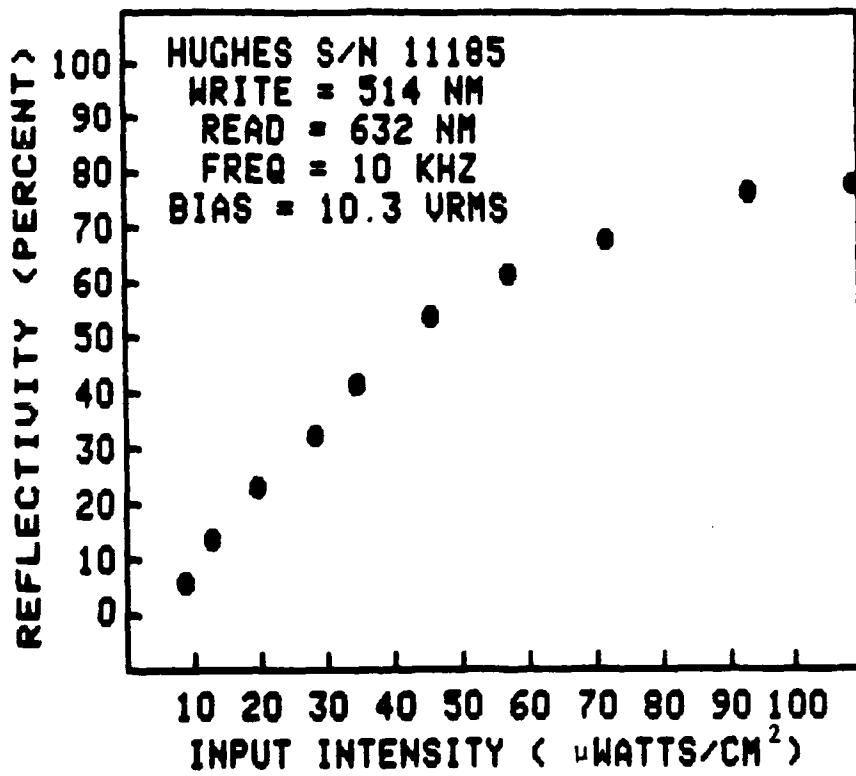


Figure 7 - Reflectivity of the Hughes LCLV at a wavelength of 632.8 nm as a function of writing light intensity at a wavelength of 514.5 nm. The AC bias voltage is 10.3 VRMS.

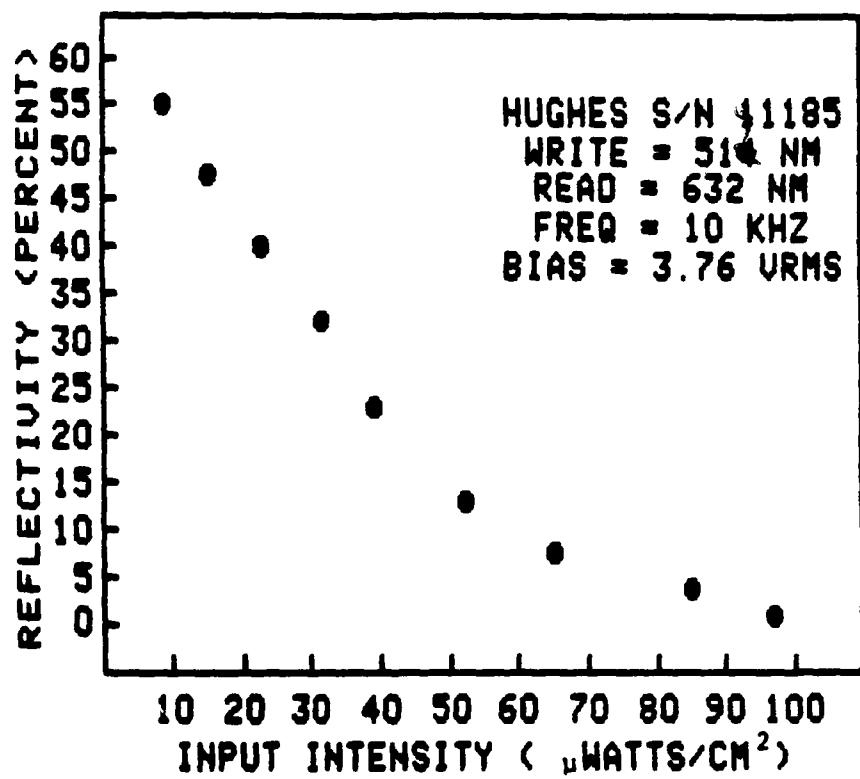


Figure 8 - Reflectivity of the Hughes LCLV at a wavelength of 632.8 nm as a function of writing light intensity at a wavelength of 514.5 nm. The AC bias voltage is 3.76 VRMS.

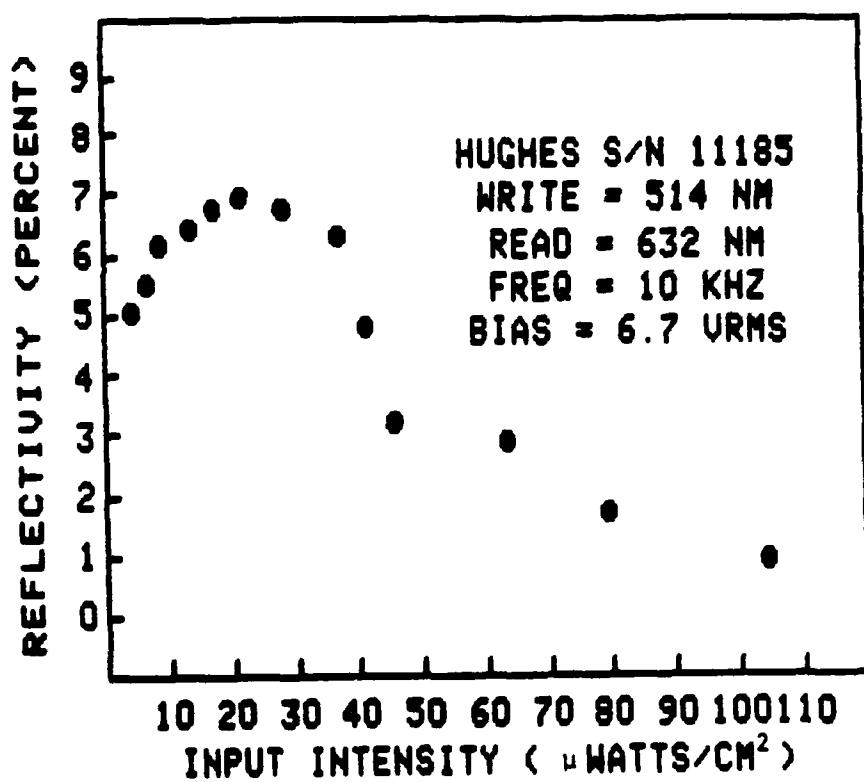


Figure 9 - Reflectivity of the Hughes LCLV at a wavelength of 632.8 nm as a function of writing light intensity at a wavelength of 514.5 nm. The AC bias voltage is 6.7 VRMS.

E. Writing Wavelength Sensitivity

As mentioned in Chapter II, the wavelength of the writing beam should be chosen to have a maximum effect on the conductivity of the photoconductive layer. Experiments were performed in which a constant intensity variable wavelength source was used to illuminate the LCLV.

White light was sent through a diffraction grating monochromator (having a bandpass of about 10 nm) and illuminated the liquid crystal light valve. A variac on the light source allowed the input intensity to be kept constant at 40 microwatts/cm². Intensities were measured using a radiometric filter on an EG&G detector system.

Figure 10 shows the relative reflectivity of the device as a function of writing wavelength and shows a very strong wavelength response characteristic of the Cadmium Sulfide photoconductor layer. This curve shows the extremely narrow spectral response of this device. However, it happens to coincide with the 514 nm emission line of the Argon ion laser. Note that the response is zero for the 488 nm Argon laser line.

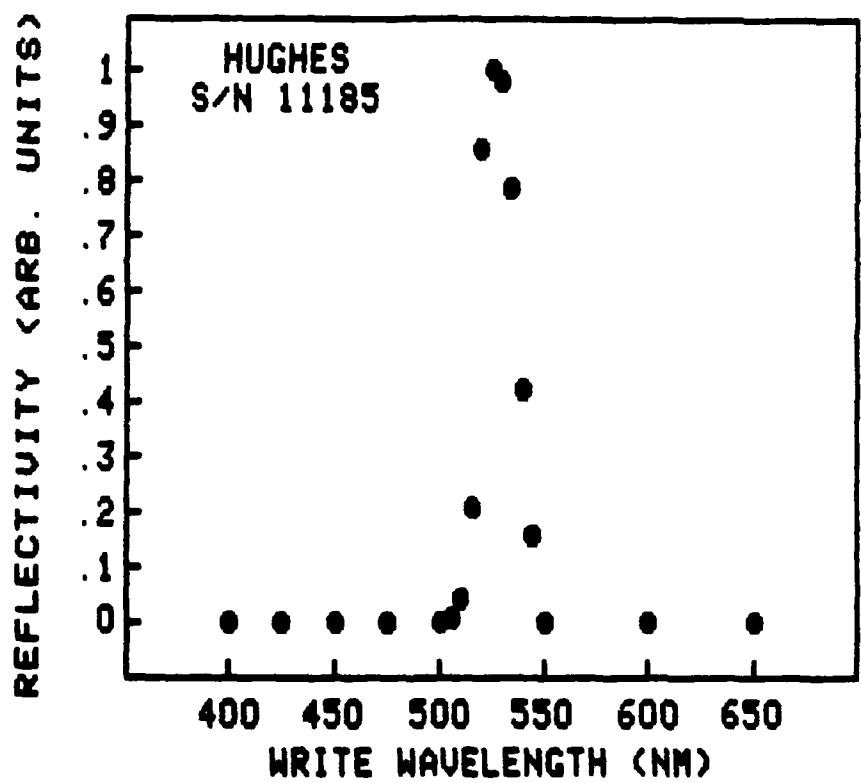


Figure 10 - Relative response of the Hughes LCLV at a wavelength of 632.8 nm as a function of writing light wavelength at constant intensity. Writing wavelength bandpass is approximately 10 nm.

F. Reading Wavelength Sensitivity

As mentioned in Chapter II, the polarization of the reading beam changes due to induced birefringence in the liquid crystal layer. The degree of birefringence not only depends on the induced orientation changes in the liquid crystal layer but also on the thickness of the liquid crystal layer and the wavelength of the reading beam. The thickness can be tuned to maximize the degree of birefringence for a chosen wavelength.

Measurements of the relative reflectivity of the light valve as a function of reading wavelength were carried out. Input white light was sent onto the light valve through a diffraction grating monochromator having a bandpass of about 10 nm.

Light reflected at a slight angle was collected by a detector. The polarization beamsplitter cube was not used because its reflectivity is not constant as a function of wavelength.

The detector was first used to measure the incident light intensity at the LCLV and then the reflected intensity. As before, the detector had a radiometric filter to give a constant response as a function of wavelength.

The results of this experiment are shown in Figure 11 and show that this device thickness is tuned to the Helium Neon laser wavelength.

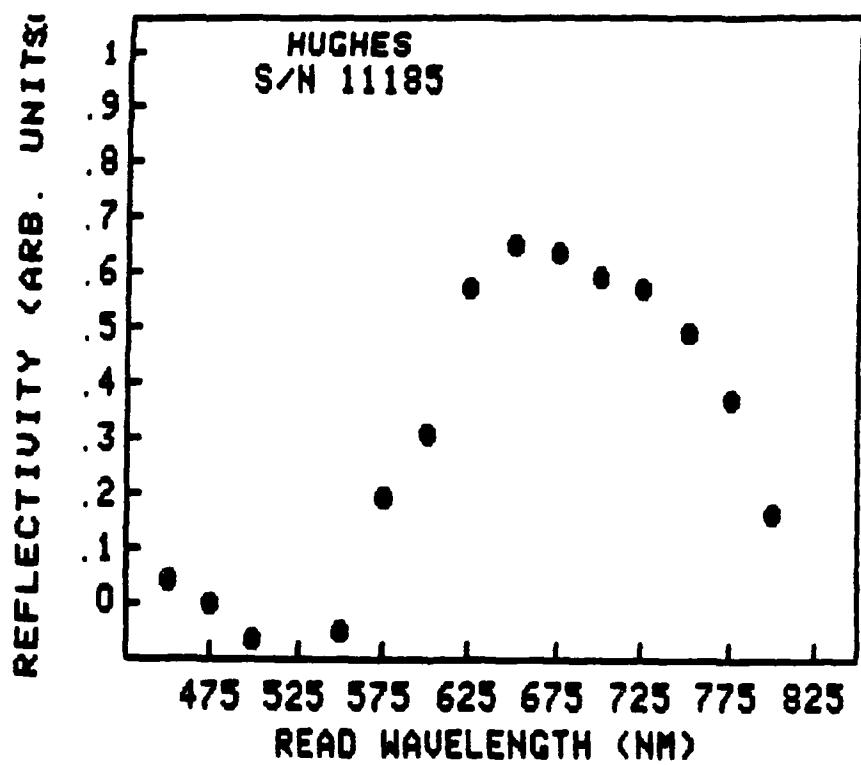


Figure 11 - Relative response of the Hughes LCLV as a function of reading light wavelength.

G. Response Time Measurements

Rise and fall times for the Hughes device were measured. In the experimental setup, a shutter mechanism was used to illuminate the LCLV with a 5 second pulse at the Argon laser line of 514.5 nm. The reflected light at a wavelength of 632.8 nm was then monitored as a function of time using a detector. System risetime and falltimes were measured by shining the writing light directly onto the detector and were substantially shorter than the LCLV response time.

Figure 12 shows the reflected signal as a function of time for different writing intensities. The bottom curve shows the system response. The three higher curves show the time dependence of the reflectivity for writing beam intensities of 11, 26, and 66.5 microwatts/cm².

The figure shows a complicating feature at the lower writing intensities. In this case, the reflectivity rises initially very quickly, then pauses, and increases much more slowly. A description of the time profile is therefore very complicated. The time required to reach 90% of the final value is extremely long while the 50% time is still quite short. We decided to take the 50% points since this feature remained qualitatively unchanged. Values for both the risetime and falltimes as a function of writing light intensity are shown in Figure 13. The falltime is relatively constant at about 35 msec. However, the risetime decreases rapidly as input writing intensity increases reaching values as low as 20 msec.

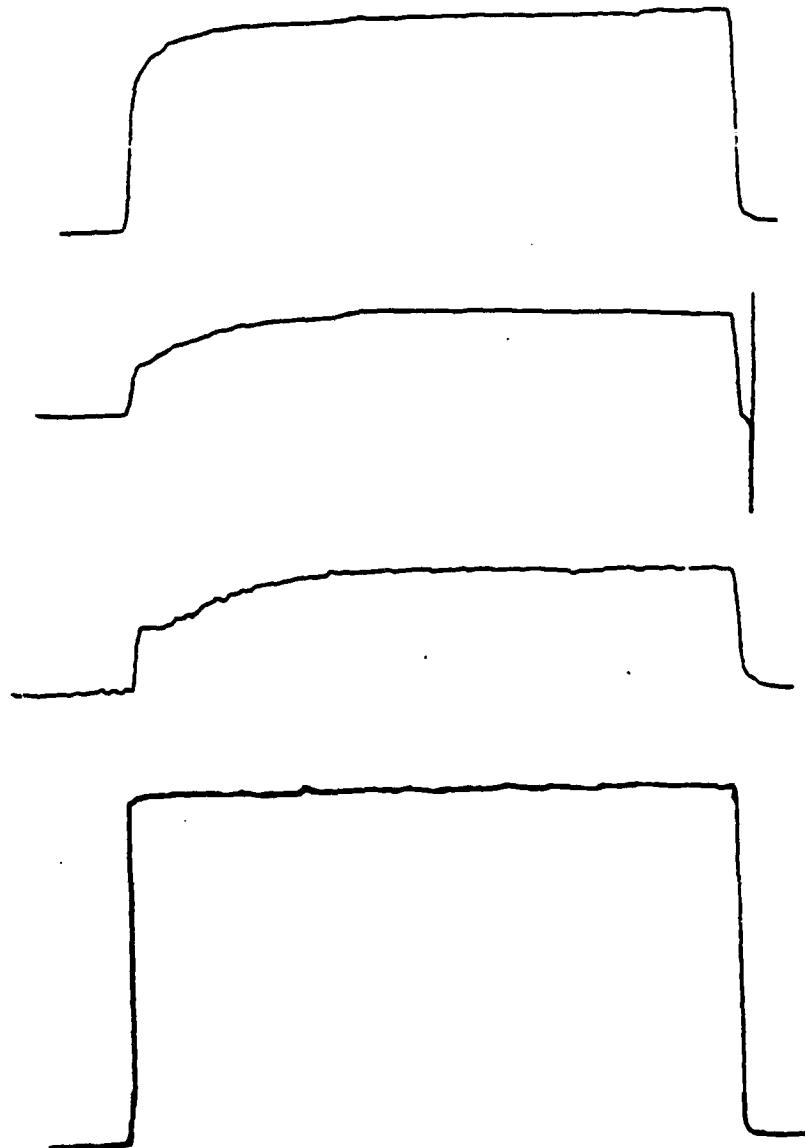


Figure 12 - Reflectivity of the Hughes LCLV at a wavelength of 632.8 nm as a function of time for different writing light intensities of (from top) 66.5, 26, and 11 microwatts/cm². The bottom trace shows the writing signal as a function of time for a 5 second pulse.

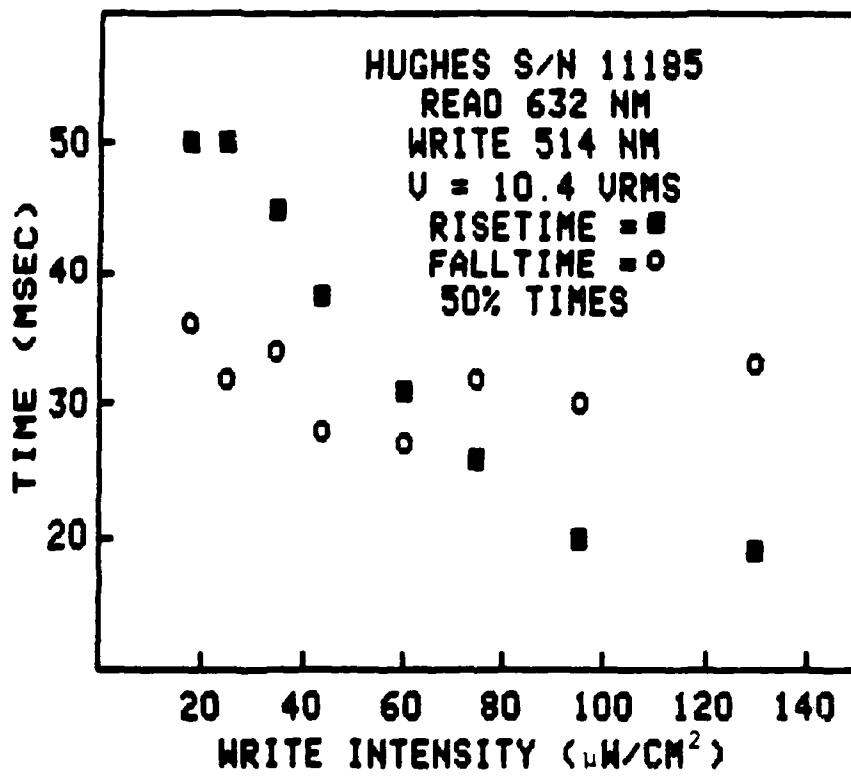


Figure 13 - Risetime and falltime of the reflectivity of the Hughes LCLV at a wavelength of 632.8 nm for different writing light intensities. The times given are the times required to reach 50% of the total response.

H. Modulation Transfer Function

A Twyman Green interferometer was used to write sinusoidal patterns on the LCLV. By changing the intensities of light from the two mirrors of the interferometer, the modulation index of the writing beam could be changed.

The read beam was reflected off the LCLV, passed through the polarization splitting cube beamsplitter, and sent through a 36.8 cm focal length Fourier transform lens to a linear diode array. The diffraction pattern consists of a large zero order peak plus two sidebands as shown in Figure 14. As the tilt of one of the mirrors in the Twyman Green interferometer increases, the spatial period of the sinusoidal pattern decreases. Therefore the location of the first order diffracted peak moves away from the zero order peak and is detected by different elements of the linear diode array. As the response of the LCLV to higher spatial frequencies decreases, the intensity of the first order peak decreases, allowing direct measurement of the MTF.

Figure 15 shows representative curves showing an overlay of the diffracted intensities in the focal plane of the lens as the tilt of the interferometer changes. The four peaks all correspond to the first order peak and clearly show the decreasing response at higher spatial frequencies.

Figure 16 shows the resulting normalized MTF curve. The spatial frequency corresponding to the 50% point is about 11 lines/mm.

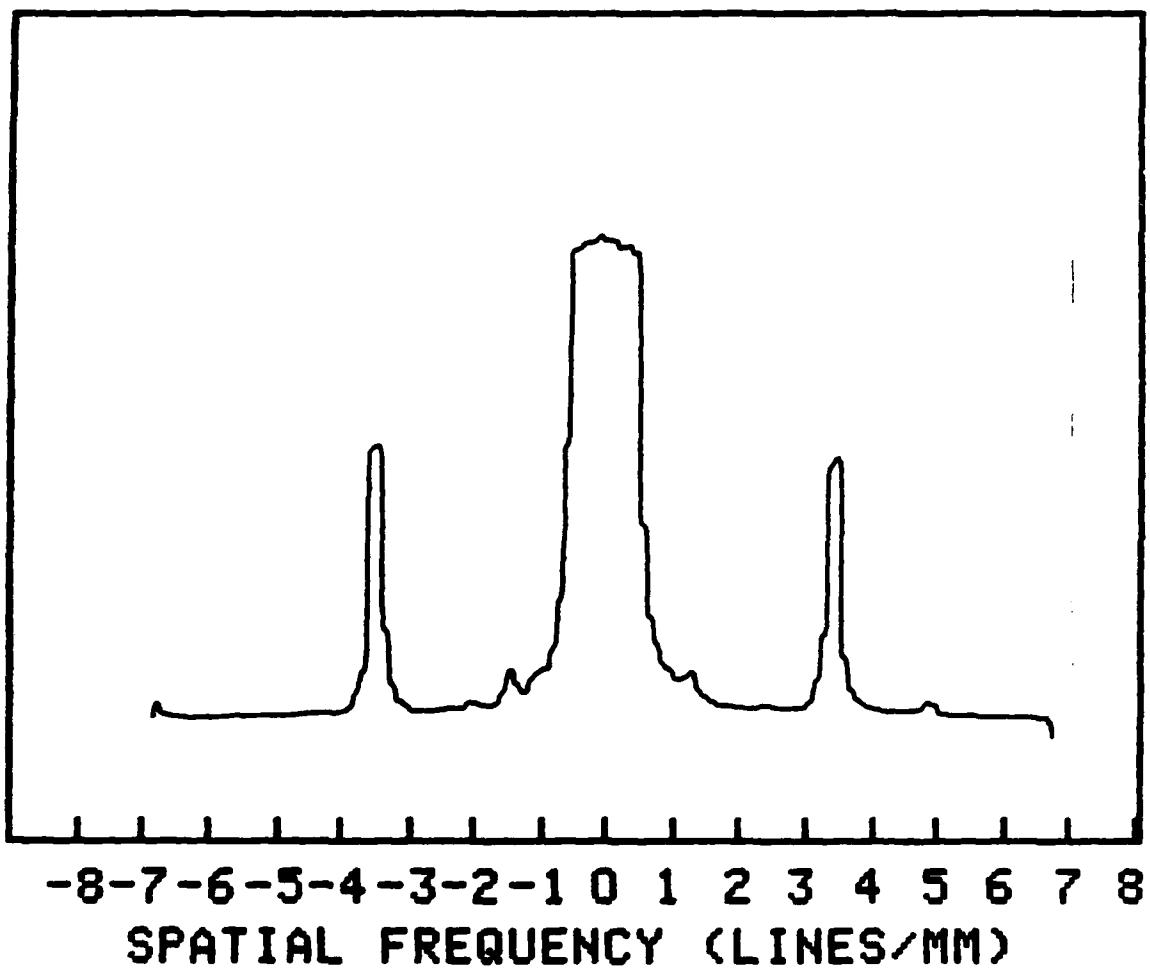


Figure 14 - Diffraction pattern of the read beam using the interferometer technique discussed in the text. The DC and two first order peaks are shown.

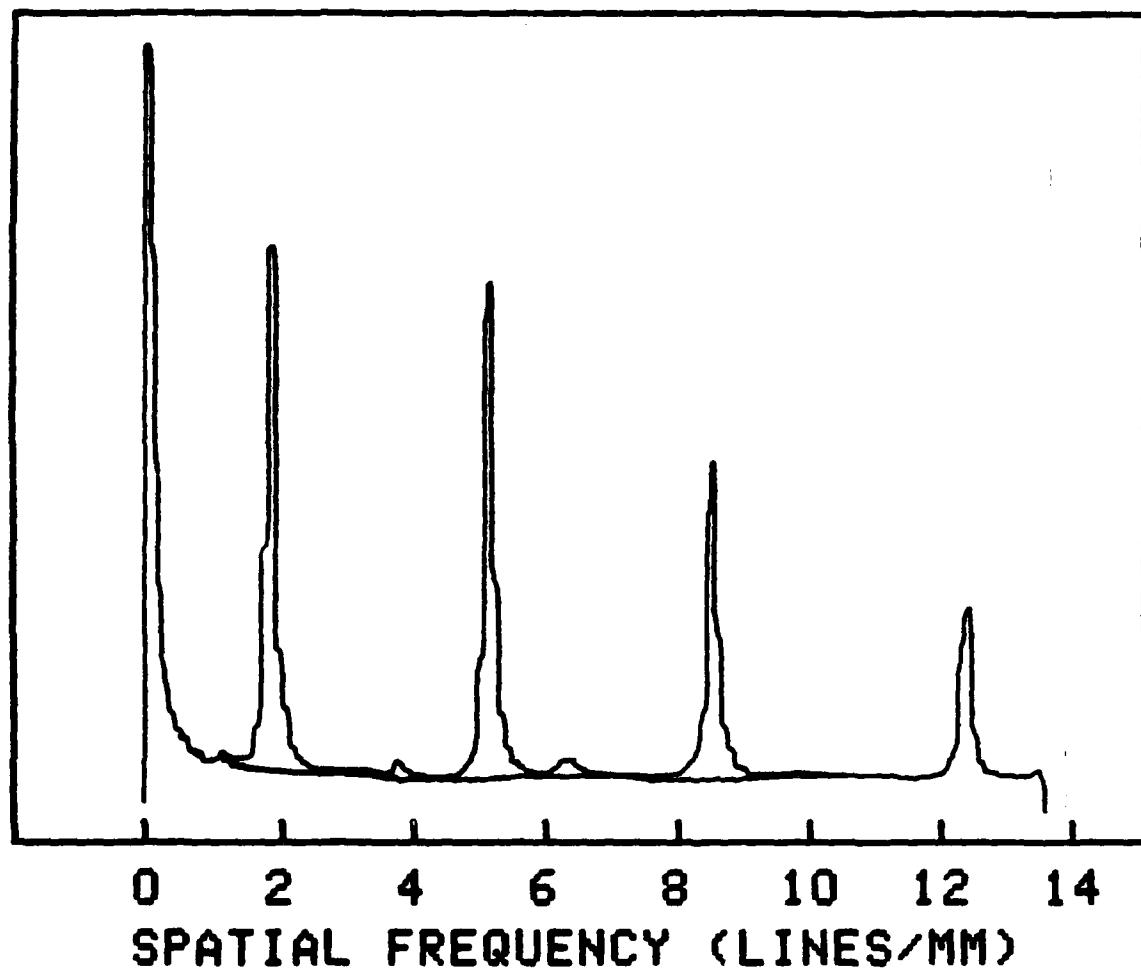


Figure 15 - Relative heights of the first order diffracted peaks of the reading beam as a function of spatial frequency in the writing beam.

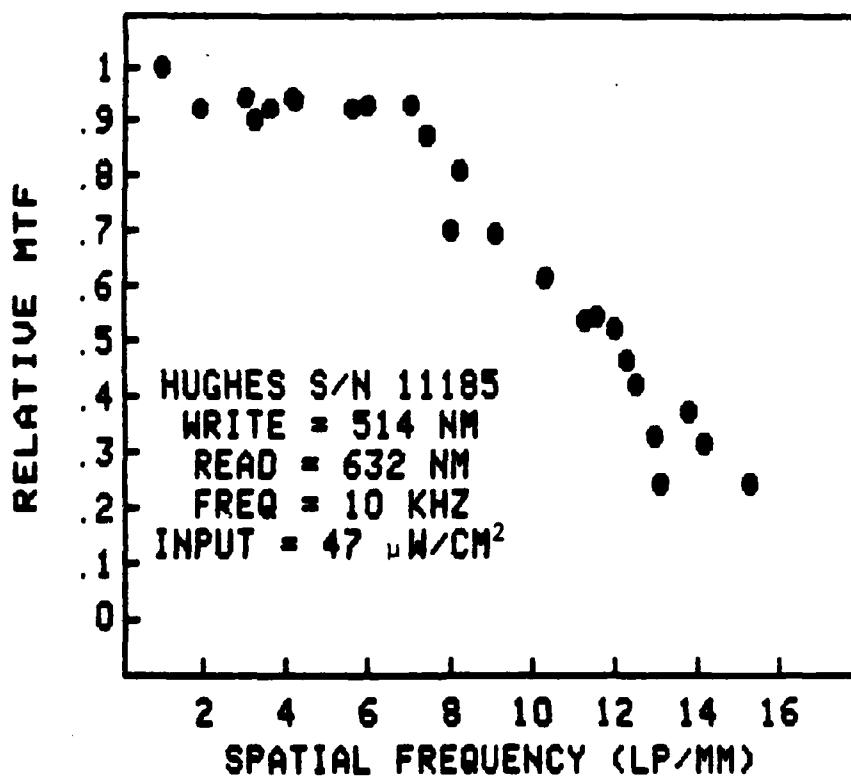


Figure 16 - MTF Curve for Hughes LCLV

I. Time Response vs. Spatial Frequency

An experiment was carried out examining the time dependence of a spatial pattern as a function of spatial frequency. Using the previous setup, a single detector replaced the detector array and was positioned to detect the first order diffracted peak. Then a shutter was used to turn on the writing beam as discussed in Section G. The time dependence of the diffracted peak was examined for different spatial frequencies in the writing beam pattern. No significant differences were observed.

J. Conclusions Regarding the Performance of the Hughes Device

The Hughes device performs extremely well. Writing beam intensities of only 50-100 microwatts/cm² are required to achieve moderate reflectivities. The maximum reflectivity of the device is about 80%.

The writing wavelength range over which the device is sensitive is unfortunately narrow and is determined by the choice of CdS as the photoconductive layer. Nevertheless, this spectral region overlaps nicely with the emission wavelengths of phosphors used in CRT and image intensifier tubes as well as the 514.5 nm emission line of the Argon laser. However, newly developed devices by Hughes use a silicon photoconductor (9) and the wavelength response is much broader.

The curve of reflectivity versus writing beam intensity can be varied by proper choice of the bias voltage. For normal operation, values near 10.3 VRMS must be chosen. However, an interesting reverse contrast mode can be obtained for a low bias voltage of 3.73 VRMS.

Response times are in the 20 - 35 millisecond range. Spatial resolution is also good with the 50% MTF point at about 11 lines/mm.

Overall, we felt that this device performed extremely well.

IV. PERFORMANCE OF THE LORAL LCLV

A. Introduction

As was mentioned earlier, the Loral device is still in development stages. When we received the device, the operating instructions were vague regarding writing sensitivity, writing wavelength, reading wavelength, and operating frequency. We experienced a great deal of difficulty in getting the device to work for two reasons.

First, the amount of writing intensity which was required is substantially greater than in the case of the Hughes device and is also much greater than previously published (6). We lost some time before discovering this.

Secondly, we lost a substantial amount of time due to a crack developing in the conducting epoxy. This prevented the bias voltage from reaching the device package. We repaired this by applying a new coat of conducting epoxy. However, this raised questions about the durability of the device.

We repeated the same series of tests on this device as on the Hughes device.

B. General Experimental Setup

The general experimental setup was conceptually the same as discussed earlier. However, we did not use the Argon laser in the initial experiments for two reasons. First, we wanted to examine the writing wavelength sensitivity. Secondly, the amount of writing light intensity was greater than we were able to obtain with our Argon laser.

The writing source was a 100 Watt tungsten light source manufactured by Oriel which was spectrally filtered to obtain a moderately monochromatic source. High and low pass optical filters having cut-off wavelengths spaced by 50 nm were combined in series to create a series of 50 nm wide spectral sources. This bandwidth was sufficient to obtain sufficient writing light intensities to conduct the experiments.

The reading light beam was a Helium Neon laser beam as before.

C. Bias Voltage

With the writing beam turned off, the absolute reflectivity of the Loral device was measured as a function of bias voltage and results are shown in Figure 17. The bias frequency was fixed at 90 Hz which is much higher than recommended by the manufacturer. The reasons for this will become clearer in future sections.

The observed characteristics are similar to those seen with the Hughes device. The reflectivity is initially high, decreases to a minimum, goes back through a small peak, decreases, and then goes through a large peak again.

In order to minimize the reflectivity when there is no writing beam (as is usually desired), a bias of 2.59 VRMS would be used. However, if maximum reflectivity is desired in the absence of any writing beam, a bias voltage of 3.48 VRMS would be used. The consequences of this have already been discussed in the case of the Hughes device and will also be noted later.

The reflectivity as a function of bias voltage is next plotted in Figure 18 for a writing beam intensity of 2.5 milliwatts/cm².

In this case, the writing beam has a 50 nm spectral width centered at 625 nm. As in the case of the Hughes device, the curve is the same except that the peaks have moved to lower voltages. As before, a larger fraction of the bias voltage appears across the liquid crystal layer because of the presence of the writing beam and the reflectivity of the device has

increased.

Therefore, if the bias voltage is chosen at 2.59 VRMS, the reflectivity will increase as the writing beam intensity increases. Next the wavelength sensitivity of the writing beam will be examined.

As in the case of the Hughes device, it is critical that the plane of polarization for the incident light be parallel to the direction of the liquid crystal molecules at the surface. We observed that slight rotations of the device shifted the bias voltage curve slightly.

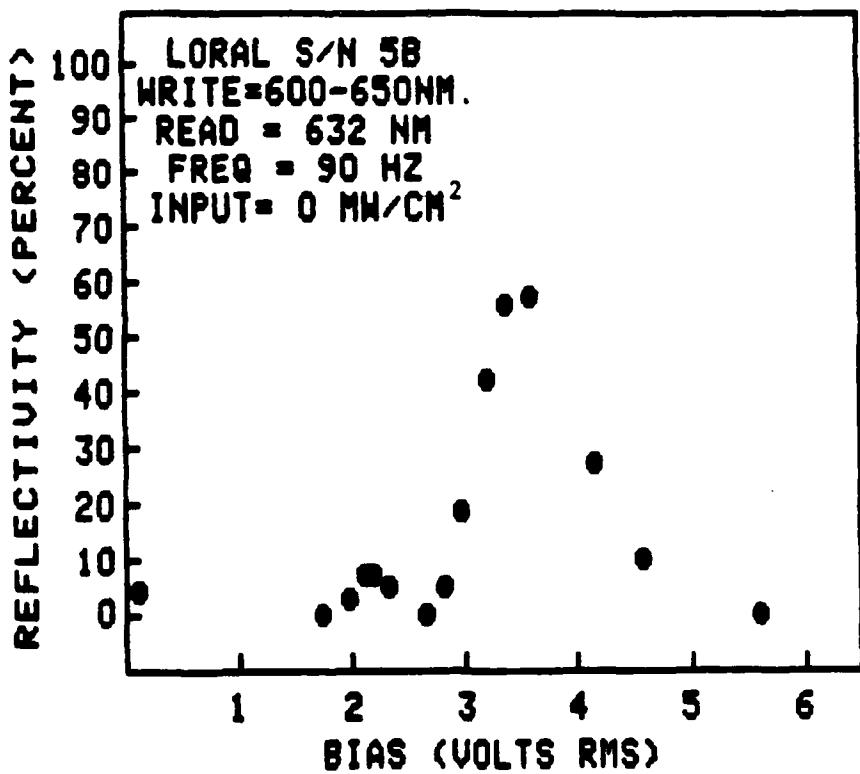


Figure 17 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of AC bias voltage applied across the device. Here, the writing beam is turned off.

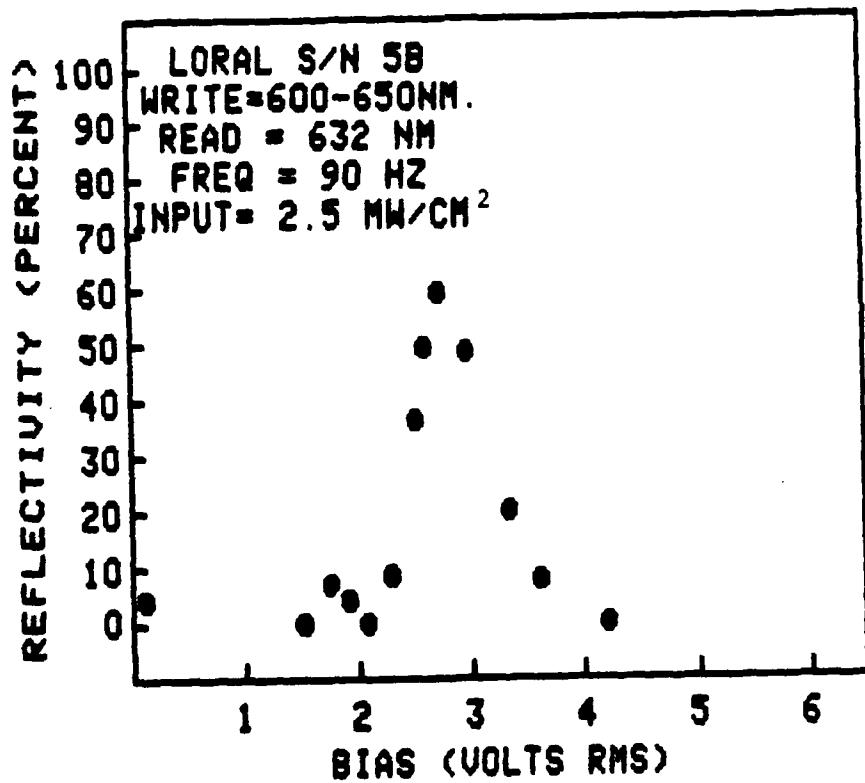


Figure 18 - Reflectivity of the Loral LCLV at a wavelength of 632.0 nm as a function of AC bias voltage applied across the device. Here, a writing beam of 2.5 milliwatts/cm² centered at 625 nm is used.

D. Writing Wavelength Sensitivity

As mentioned in the introduction, the wavelength of the writing beam should be chosen to have a maximum effect on the change in the photoconductive layer. Since we had no experimental data on proper writing wavelengths, we decided to perform this experiment before conducting sensitivity experiments. Figure 19 shows the relative reflectivity of the device as a function of the writing wavelength and shows a very strong wavelength response characteristic of the As_2S_3 photoconductor layer.

In this experiment, the writing light was spectrally filtered using high and low pass interference filters and the intensity was kept constant at 0.8 milliwatts/cm². This experiment shows that the spectral region near 625 nm is the optimum writing wavelength region and that the Argon laser is not well suited for this device.

However, the Helium Neon laser wavelength is very well suited for this device. However as mentioned earlier, the absence of any light blocking layer will cause some leakage of this light into the reading side of the LCLV.

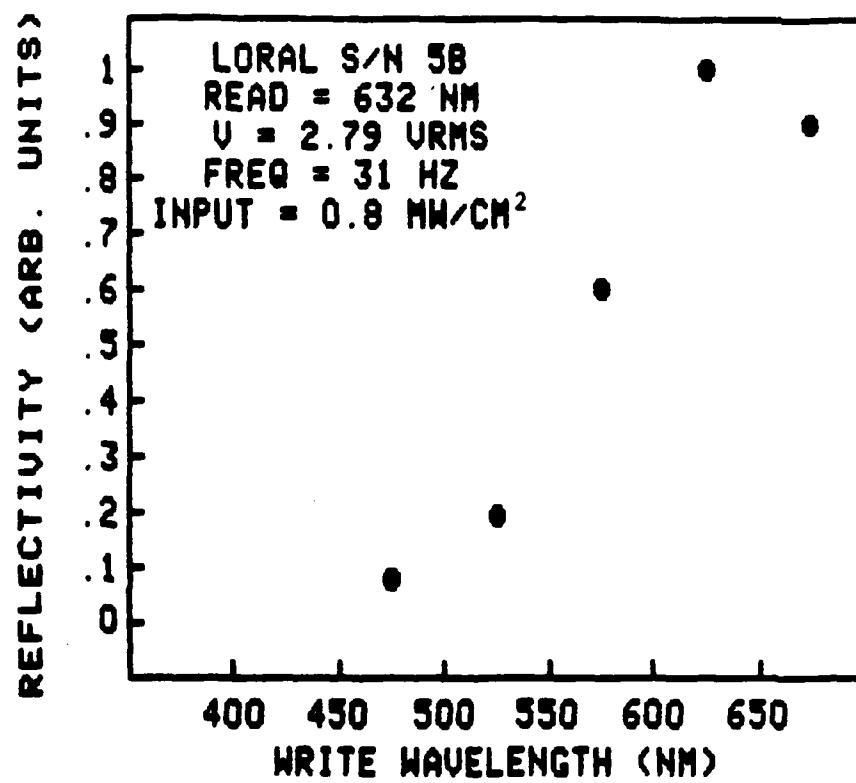


Figure 19- Relative response of the Loral LCLV as a function of writing light wavelength at a constant writing intensity.

E. Writing Sensitivity

Once the optimum writing wavelength was found, the absolute reflectivity of the Loral device at a reading wavelength of 632.8 nm was measured as a function of the intensity of the writing beam intensity. Here the spectral composition of the writing beam corresponded to the most sensitive wavelength response region and had a bandwidth of 50 nm centered at 625 nm.

Figure 20 shows the device reflectivity as a function of writing intensity. Since the device bias was chosen to be 2.59 VRMS, the reflectivity increased with increasing writing intensity. Note that moderately high intensities of 4 milliwatts/cm² are needed to achieve moderate reflectivities of about 60%. These intensities are much higher than in the case of the Hughes device.

A sensitivity measurement using the Helium Neon laser as the writing source gave identical results.

An interesting effect takes place is the bias voltage is chosen at 3.48 VRMS. Recall that in this case, the reflectivity was a maximum when the writing intensity was zero. Now as the intensity of the writing beam increases, the curve shown in Figure 17 again moves to the left causing the reflectivity to decrease.

This is shown in Figure 21 and shows the reflectivity decreasing as the writing beam intensity increases.

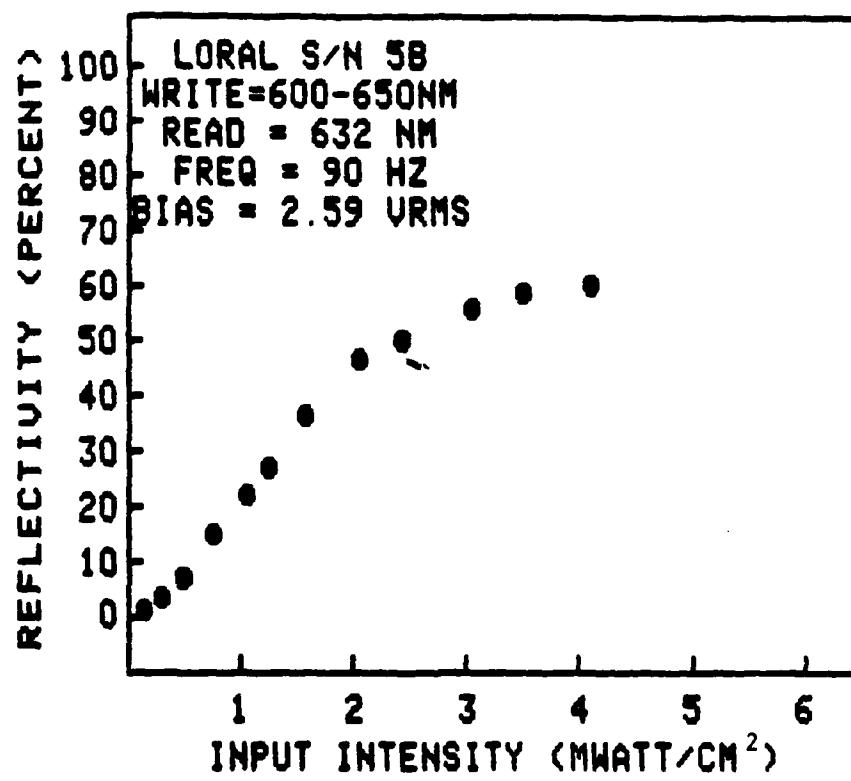


Figure 20 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of writing light intensity at a wavelength of 625 nm. The AC bias voltage is 2.59 VRMS.

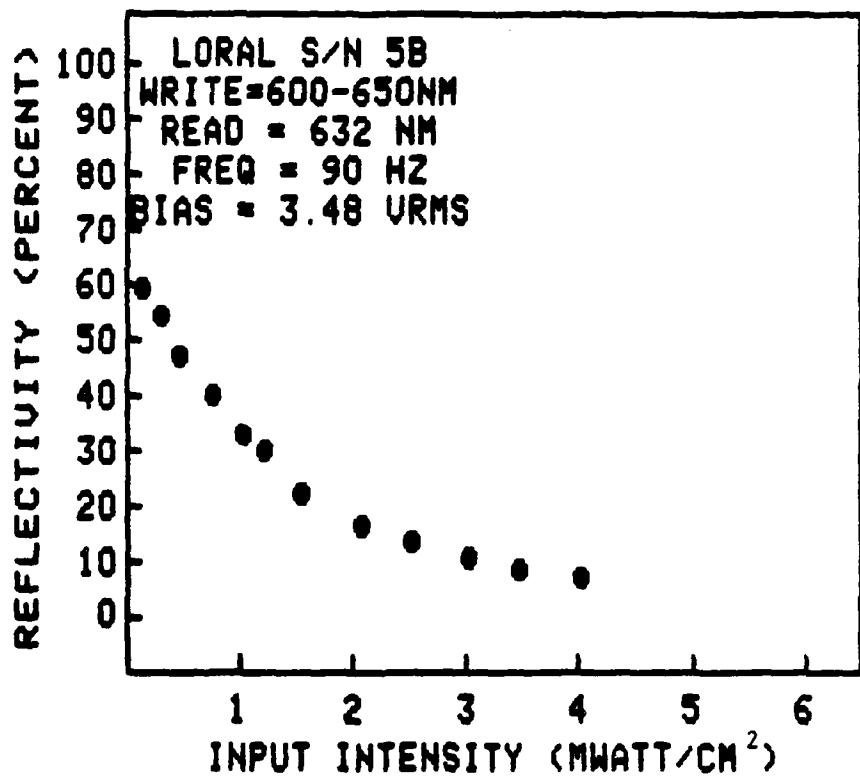


Figure 21 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of writing light intensity at a wavelength of 625 nm. The AC bias voltage is 3.48 VRMS.

F. Reading Wavelength Sensitivity

As mentioned in Chapter II, the polarization state of the reading beam changes due to induced birefringence in the liquid crystal layer. The degree of birefringence not only depends on the induced changes in the liquid crystal layer but also on the thickness of the liquid crystal layer and the wavelength of the reading beam. The thickness can be tuned to maximize the degree of birefringence for a chosen wavelength.

Measurements of the reflectivity of the light valve as a function of reading wavelength were carried out. The experimental setup was the same as with the Hughes device. Input light was sent onto the light valve through a diffraction grating monochromator slightly off normal incidence. The reflected light was collected by a detector. The polarization beamsplitter cube was not used because of its wavelength response.

The detector measured both the incident light intensity at the LCLV and then the reflected intensity. As before, the detector had a radiometric filter to give a constant response as a function of wavelength.

The results of this are shown in Figure 22 and show that this device thickness is tuned to the Helium Neon laser wavelength. However, note from the earlier discussion that the lack of a blocking layer would allow a portion of the reading light at these wavelengths to be transmitted to the photoconductor and act as a writing beam.

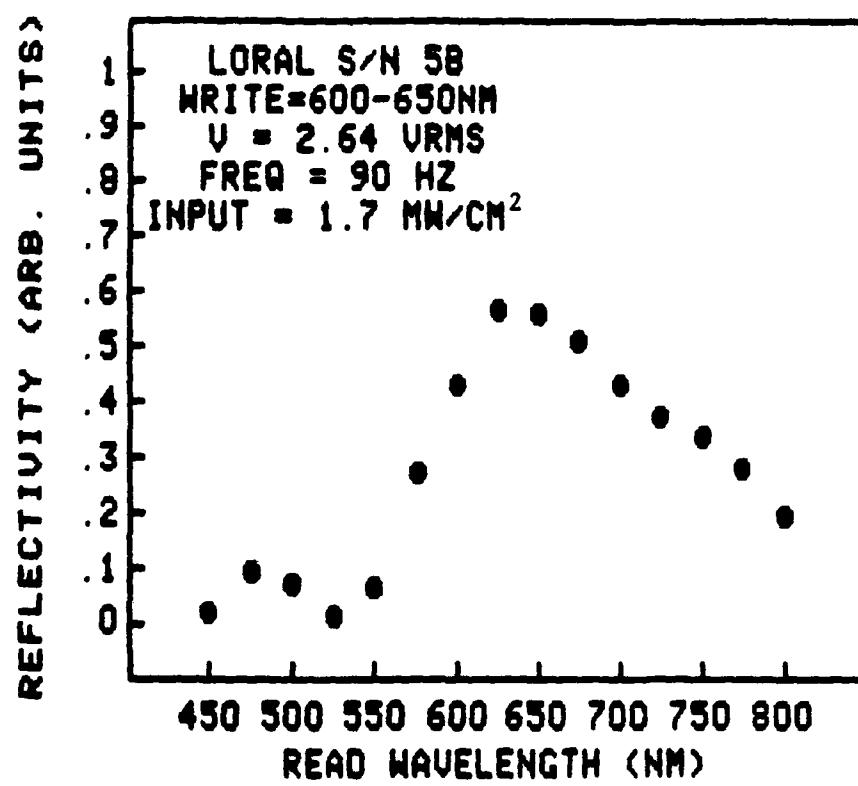


Figure 22 - Relative response of the Loral LCLV as a function of reading light wavelength.

G. Bias Voltage Frequency Response

The instructions for the Loral device suggested using bias frequencies of only 10 - 40 Hz. This is much lower than in the case of the Hughes device. We noted, when using a 1 Hz bias frequency, that the reflected intensity would flicker. We became concerned with this behavior since the flicker would be indistinguishable from changes in the incident light intensity level.

To check this, we performed tests with a transient writing intensity for a variety of bias frequencies and a given bias voltage. This not only enabled us to study the dependence on the reflectivity with bias frequency, but also permitted a study of the risetime characteristics of the device.

As in the related experiment which we did on the Hughes device, the writing light was sent through a shutter with an exposure time of 2 seconds. The reflected signal was again sent to a fast detector and recorded. The system risetime was much faster than the time response of the LCLV.

Figure 23 shows the reflected signal for a bias frequency of 20 Hz. Not only is the risetime extremely slow, but there is a substantial ripple since the liquid crystal molecules are reacting to the changing bias voltage. The ripple frequency follows the bias frequency.

Figures 24-26 show the experimental results at bias frequencies of 60, 90, and 160 Hz. We see that the absolute reflectivity passes through a maximum and then decreases with

increasing bias frequency. In addition, the ripple decreases with increasing bias frequency while its frequency increases. Figures 27 and 28 plot the relative response and percent ripple as a function of bias frequency. These curves show that the operating condition which maximizes the response while simultaneously minimizing the ripple occurs at a bias frequency of 90 Hz. This was the value usually used in the experiments reported earlier.

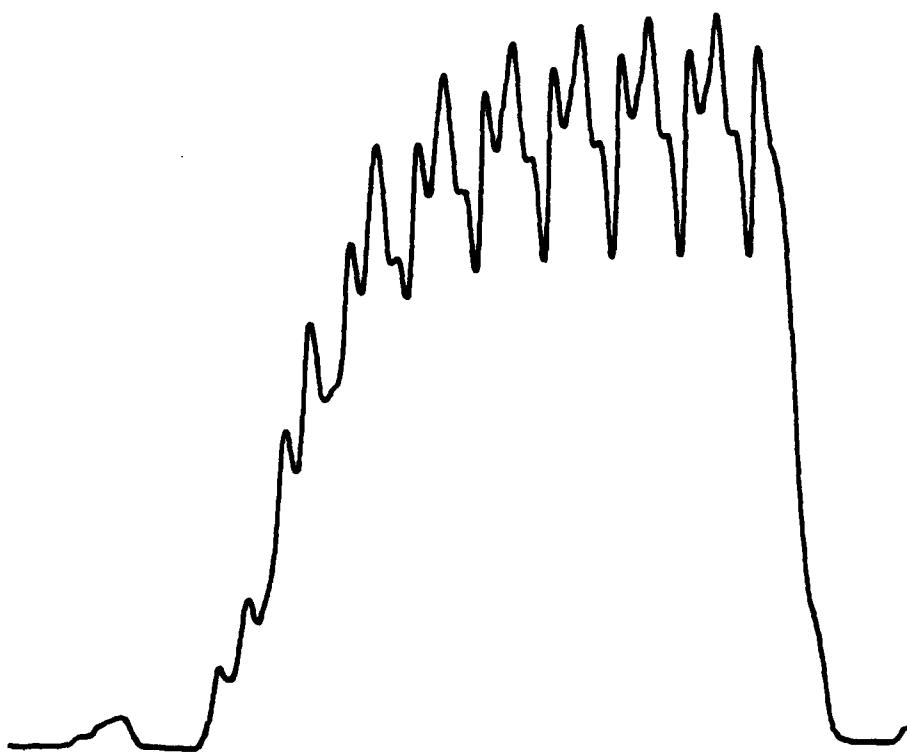


Figure 23 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of time for a bias voltage frequency of 20 Hz.

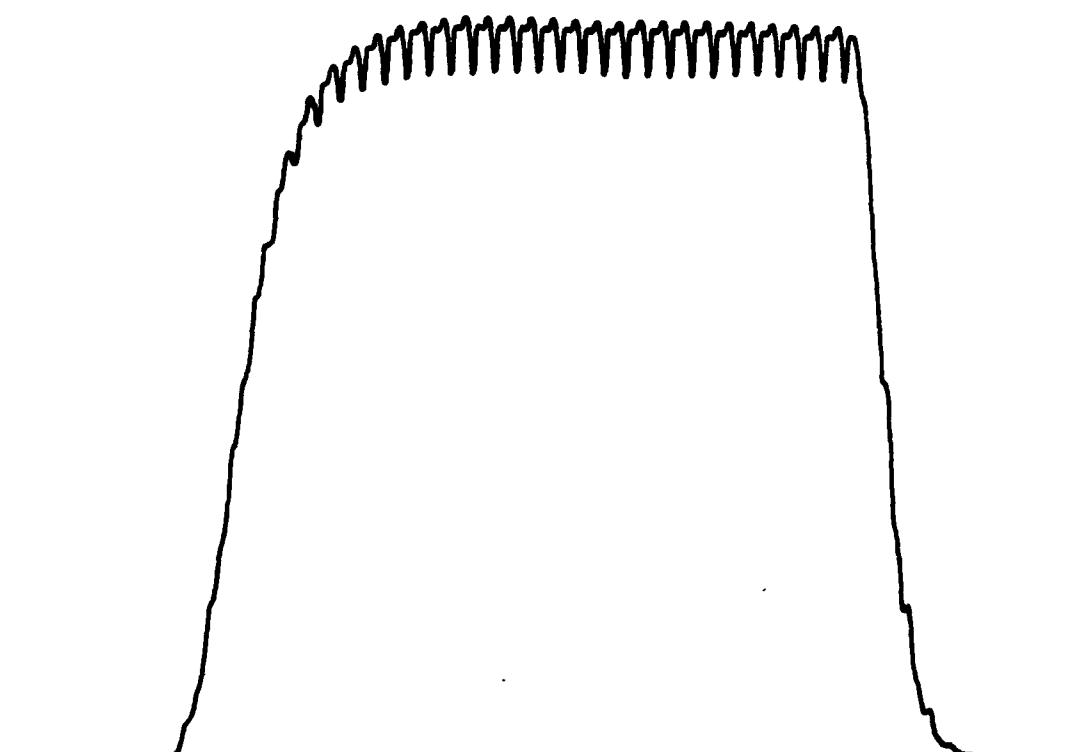


Figure 24 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of time for a bias voltage frequency of 60 Hz.

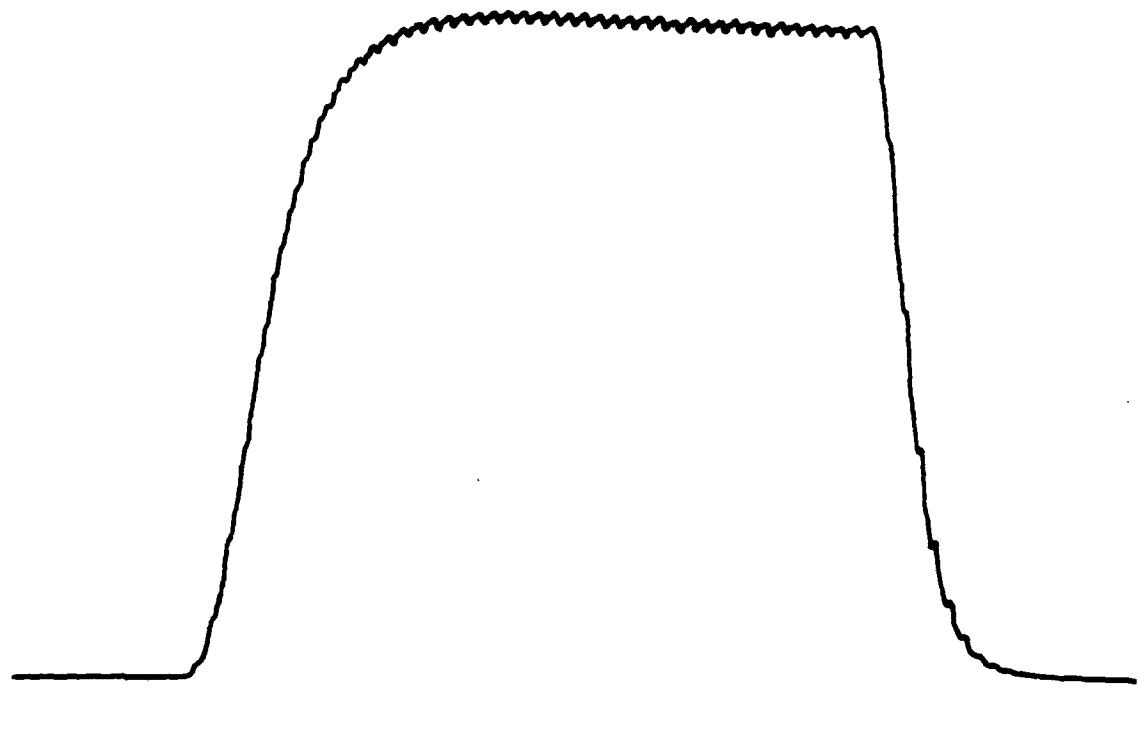


Figure 29 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of time for a bias voltage frequency of 90 Hz.



Figure 26 - Reflectivity of the Loral LCLV at a wavelength of 632.8 nm as a function of time for a bias voltage frequency of 160 Hz.

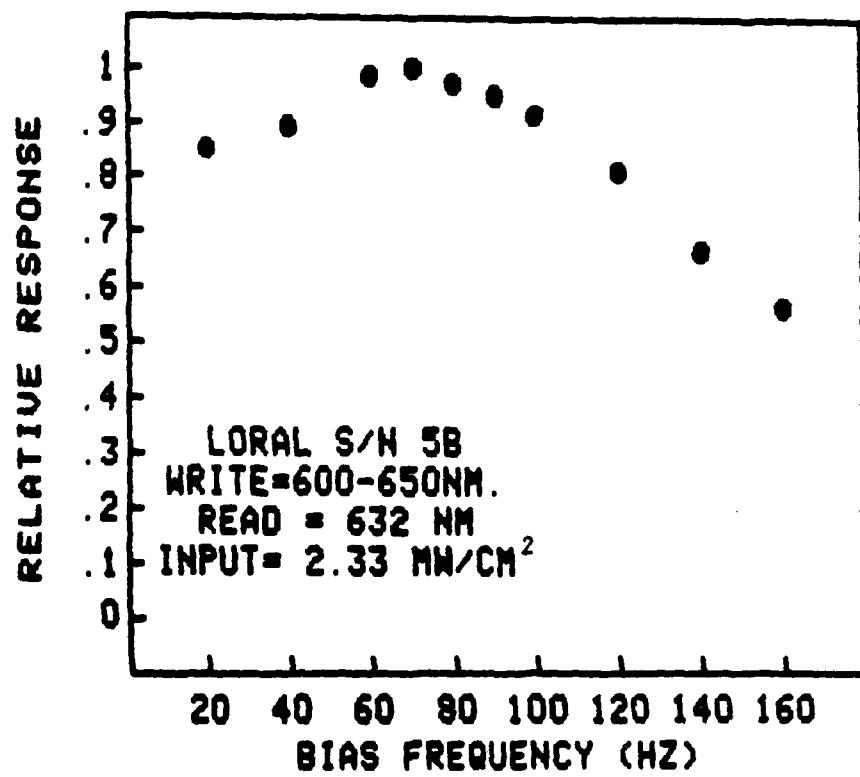


Figure 27 - Relative response of the Loral LCLV as a function of bias voltage frequency.

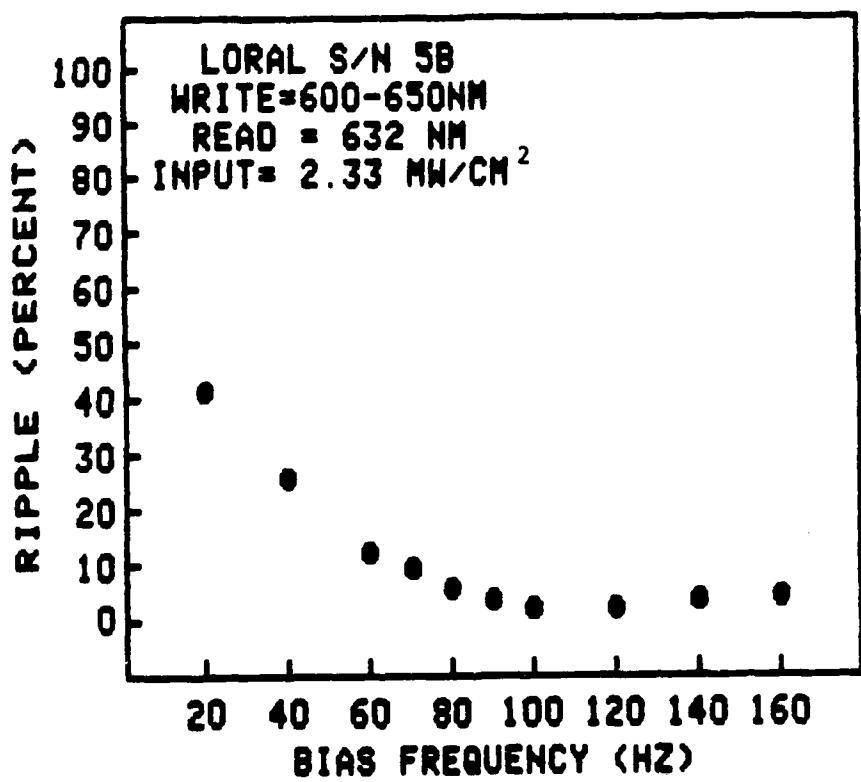


Figure 28 - Percent ripple for the Loral LCLV as a function of bias voltage frequency.

H. Response Time

Rise and fall times for the Loral device were measured using the same experimental setup as in the previous section. However, we made measurements of their dependence on both the bias frequency and on writing intensity.

Figure 29 shows that the risetime decreases with increasing bias frequency and flattens out above 60 Hz.

Figure 30 shows that, using a bias frequency of 90 Hz, the risetime and falltime are relatively independent of writing intensity and are about 50 and 20 milliseconds respectively. These numbers compare well with results using the Hughes device.

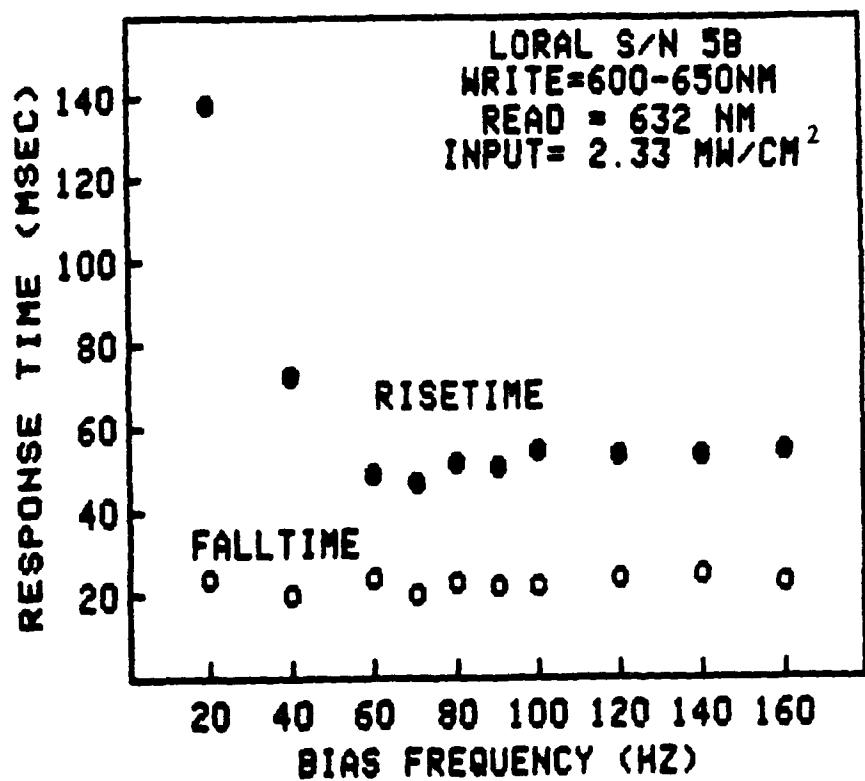


Figure 29 - Response time of the Loral LCLV as a function of bias frequency.

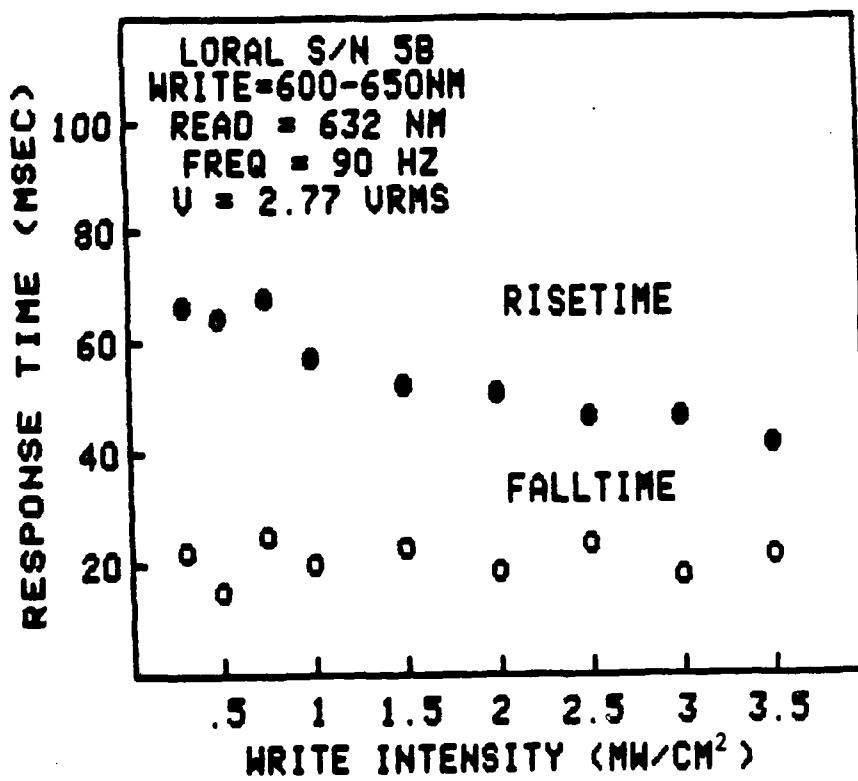


Figure 30 - Response time of the Loral LCLV as a function of writing light intensity.

I. Modulation Transfer Function

As in the case of the Hughes device, a Twyman Green interferometer was used to write sinusoidal patterns on the LCLV. However, we were not able to obtain any reportable experimental results due to the fact that the Argon laser is not well suited for writing on the Loral device. As reported earlier, the photoconductor used in the Loral device is quite insensitive to the Argon laser wavelength. After spatially filtering, expanding, and aperturing the Argon laser beam, we were unable to obtain sufficient writing intensities. In fact, we measured a reflectivity of only 5% with an incident intensity at 514.5 nm of 2 milliwatts/cm². When we removed the spatial filter and changed the collimating optics to obtain a higher writing intensity, the writing beam uniformity was insufficient and degraded the MTF of the writing system. Therefore, this interferometric technique could not be carried out.

J. Conclusions Regarding the Sensitivity of the LORAL Device

As mentioned earlier, the Loral device is still in development stages and suffers from a lack of detailed operating instructions. We found that the recommendations given by the manufacturer were incomplete and often incorrect. In addition, we had troubles with the structural ruggedness of the device.

The chief drawback to this device is that it is substantially less sensitive than the Hughes device. Writing beam intensities of 2 - 3 milliwatts/cm² are required to achieve moderate reflectivities. The maximum reflectivity of the device is about 60% which is good.

The wavelength range over which the device is sensitive is much broader than in the case of the Hughes device but is centered in the red spectral region. Therefore it offers a good match with several red and near-infrared laser sources.

The lack of the light blocking layer presents a serious obstacle since many common diode laser sources, which might be used in correlators, operate at these wavelengths.

As in the case of the Hughes device, the curve of reflectivity versus writing beam intensity can be varied by proper choice of the bias voltage. For normal operation, values near 2.59 VRMS must be chosen. However, an interesting reverse contrast mode can be obtained for a bias voltage of 3.48 VRMS.

We found substantial variation in the device performance as a function of bias voltage frequency which affected the absolute response, ripple, and time response. We recommend a bias voltage

frequency of 90 Hz which is much higher than recommended by the manufacturer.

The difference in the operating frequencies for the Loral and Hughes devices can be attributed to the different equivalent circuits attributed to the various constituent layers including the photoconductor, the mirror, the liquid crystal, and the optical blocking layer. This effective circuit which includes the diode formed by the photoconductor/mirror layer is driven by the AC voltage and has an optimum driving frequency. It is impossible to predict this optimum frequency without separately testing each of the components. Neither Loral nor ourselves was willing to tear apart the devices to do this. Therefore, we have only reported the total frequency response.

The response time of the device was certainly adequate once the proper operating frequency was obtained.

The insensitivity of the device in the green spectral region prevented our making MTF measurements.

The final section will offer a few direct comparisons and conclusions.

V. DISCUSSION

Based on the measurements reported in the last two chapters, we recommend using the current version of the Hughes device over the current version of the Loral device for use in an optical correlator at this time.

First, in comparing the sensitivity at the optimum wavelength for each device, the Hughes LCLV is more sensitive by a factor of 30. The Hughes LCLV requires writing intensities of only 50-100 microwatts/cm² to achieve moderate reflectivities. The maximum reflectivity of the Hughes LCLV is about 80%. By contrast, the Loral LCLV requires writing beam intensities of 2 - 3 milliwatts/cm². The maximum reflectivity of the device is about 60%. In reporting this, we note a still unexplained discrepancy between our measurements of the writing powers needed for the Loral device with the earlier reported values (6).

Secondly, although the writing wavelength sensitivity for the Hughes device is extremely narrow, this spectral region overlaps nicely with the emission wavelengths of phosphors used in CRT and image intensifier tubes as well as with the 514.5 nm emission line of the Argon laser. However, a newly developed LCLV by Hughes uses a silicon photoconductor (9) and the wavelength response is much broader. By contrast, the spectral response of the Loral LCLV is much broader than in the case of the Hughes LCLV but is centered in the red spectral region. Therefore, it does not match the phosphor wavelengths mentioned

above.

Both devices have a broad reading wavelength sensitivity which has been maximized for the red spectral region. This is fortunate since most correlators would use diode laser sources for the reading beams. However, the lack of the light blocking layer in the Loral LCLV presents a serious obstacle since a portion of the reading beam would pass through the light blocking layer and serve as a writing beam. This problem is compounded since the writing wavelength sensitivity curve for the Loral LCLV is peaked in this red spectral region. Therefore, we would recommend inclusion of this light blocking layer into the Loral LCLV.

Both devices operated in similar fashions. The curve of reflectivity versus writing beam intensity can be varied by proper choice of the bias voltage. In both cases, the devices could be operated with increasing reflectivity for increasing writing intensity or in a reverse mode.

The Loral LCLV has an advantage in that its operating frequency and operating AC bias voltage are substantially lower than the Hughes device. We found substantial variation in the Loral LCLV performance as a function of bias voltage frequency which affected the absolute response, ripple, and time response. We recommend a bias voltage frequency of 90 Hz which is much higher than recommended by the manufacturer.

By following our recommendations and raising the operating frequency to the 90 Hz range, serious ripple in the reflectivity

as a function of time can be avoided without sacrificing responsivity.

Response times for both devices are in the 20 - 60 msec range and are adequate for most applications. The Hughes LCLV was slightly faster.

Spatial resolution is also good in the case of the Hughes LCLV with the 50% MTF point at about 11 lines/mm. We were unable to make the MTF measurement on the Loral LCLV because of its insensitivity in the green spectral region at which our argon laser operates.

Finally, the Loral LCLV is still in development stages compared to the Hughes LCLV. We feel that the structural ruggedness of the Loral LCLV must be substantially improved. In addition, the Loral LCLV suffers from a lack of detailed operating instructions. We found that the recommendations given by the manufacturer were incomplete and often incorrect. In spite of this, we feel that we have successfully examined a number of key performance characteristics of the Loral LCLV. The results of these measurements have allowed us to improve the performance of the Loral LCLV.

In conclusion, we would recommend using the Hughes device for two reasons. First, it has dramatically superior writing sensitivity. Secondly, its reading wavelength sensitivity matches the wavelength emission of common phosphors used in image intensifiers and CRTs.

VI. REFERENCES

1. J.G. Duthie, J. Upatnieks, C.R. Christensen, R.D. McKenzie, Jr, Proceedings SPIE, 231, 281, (1980)
2. J. Upatnieks, Applied Optics, 22, 2798, (1983)
3. J.G. Duthie, J. Upatnieks, Opt. Eng., 23, 007, (1984)
4. J. Grinberg et al., Optical Engineering, 14, 217, (1975)
5. W.P. Bishay et al, Optical Engineering, 17, 371, (1978)
6. W.E.L. Haas et al, Applied Physics Letters, 29, 631, (1976)
7. G. Baur in "The Physics and Chemistry of Liquid Crystal Devices", G.J Srokel ed., (Plenum Press, New York, 1980), pg. 61-78.
8. J. Grinberg et al., IEEE Trans. on Electron Devices., ED-22, 775, (1975)
9. U. Efron et al, J. Appl. Phys., 52, 1356, (1985)

APPENDIX D

**Selected
Publications Resulting from
This Contract**

PREPRINTS AND REPRINTS OF PAPERS

Report Reference Number		Page Number
23.	Tien-Hsin Chao, Hua-Kuang Liu, "Real-Time Optical Edge Enhancement using a Hughes Liquid Crystal Light Valve." OSA Spatial Light Modulator Topical Meeting (1988).	271
24.	C. Ward, H. Lamela-Rivera, H. K. Liu, "High-Gamma Spatial Light Modulator for Nonlinear Optical Processing, "SPIE 613 (1986).	275
25.	Hua-Kuang Liu, Jeffrey A. Davis, Roger A. Lilly, "Optical-Data-Processing Properties of a Liquid-Crystal Television Spatial Lights Modulator," Opt. Lett. 10, 12 (1985).	281
26.	Hua-Kuang Liu, Tien-Hsin Chao, "On the Progress of the Liquid Crystal Television Spatial Light Modulator," OSA Spatial Light Modulator Topical Meeting (1988).	287
27.	Fai Mok, Joseph Diep, Hua-Kuang Liu, Demetri Psaltis, "Real-time Computer-Generated Hologram by Means of Liquid-Crystal Television Spatial Light Modulator," Opt. Lett. 11, 11 (1986).	293
28.	Tien-Hsin Chao, Hua-Kuang Liu, "Optical Pseudocolor Image Enhancement with Real-Time Large Screen Display " Opt. Eng. 27(5), (1988).	299
29.	S. Y. Kung, H. K. Liu, "An Optical Inner-Product Array Processor for As sociative Retrieval," OE-LASE, Nonlinear Optics, Los Angeles, CA, January 19-24 (1986).	305
30.	Hua-Kuang Liu, S. Y. Kung, Jeffrey A. Davis, "Real-Time Optical Associative Retrieval Technique," Opt. Eng. 25(7), (1986).	311
31.	Hua-Kuang Liu, Tien-Hsin Chao, Jacob Barhen, Graf Bittner, "Inner-Product Optical Neural Processing and Supervised Learning," IEEE First Neural Networks Meeting, Boston, MA, Sept. (1988).	315
32.	Hua-Kuang Liu, Li-Jen Cheng, "Infrared Predetection Dynamic Range Com pression via Photorefractive Crystal," Appl. Opt. 27(6), (1988).	317
33.	Hua-Kuang Liu, Tien-Hsin Chao, "Optical Image Subtraction Techniques 1975-1985", "Hybrid Image Processing", SPIE 638, (1986).	319

Real-Time Optical Edge Enhancement Using a Hughes Liquid Crystal Light Valve

Tien-Hsin Chao and Hua-Kuang Liu
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

Edge enhancement is one of the most important preprocessing techniques utilized in optical pattern recognition. In an optical correlator, cross-correlations among similar input objects can be greatly reduced by using the edge enhancement technique. Traditionally, optical edge enhancement is obtained by high-pass filtering at the Fourier plane. However, the system SNR is generally lowered by this filtering process. Recently, two differentiating spatial light modulators, specifically designed to generate edge-enhanced output, have been reported. Casasent et al. have demonstrated real-time edge enhancement using a Priz light modulator [1]. The Priz light modulator is a transverse modification of the Pockel's Read-out Optical Modulator (PROM) with a [111] BSO crystal cut. Armitage and Thackara have designed a BSO photo-addressed nematic liquid crystal differentiating spatial light modulator [2, 3]. A layer of liquid crystal is tuned in a transverse configuration (i.e. the electro-optic response is optimized for E_x and E_y rather than E_z) to achieve the edge-enhancement. The BSO crystal is used as a photo-addressing medium. This SLM is functionally optimized as an edge-enhancing SLM.

In one of our recent artificial neural network experiments, we discovered an interesting edge enhancement effect by using a Cd-S Hughes Liquid Crystal Light Valve.

In our experiment, the edge-enhancing effect started to appear as the biasing frequency was lowered. This effect was optimized at a bias frequency of 500 Hz and voltage of 6 V_{rms} . (For normal operation, the values are 10 KHz and 10 V_{rms}). The input writing light intensity was about 50 $\mu w/cm^2$. Experimentally, a linearly polarized read-out beam was employed. First, a continuous tone image was obtained, through using a polarizing beam splitter, at the above stated biasing condition. The edge-enhanced output was obtained when the LCTV was rotated about 30° ccw.

The interpretation of this edge enhancement phenomenon is as follows: It was generally understood that in a photo-addressed SLM, the transverse field components are proportional to the spatial gradient of the writing image. Thus, a transverse SLM acts as a differentiating or edge-enhancing device [3]. As a low frequency biasing signal is applied to the Hughes LCLV, the liquid crystal molecules align themselves along the direction of the applied field that is perpendicular to the electrode surfaces. The LCLV is thus switched to a transverse mode under this bias condition and functions as an edge-enhancing spatial light modulator.

The experimental results are shown in Fig. 1. Fig. 1 (a) shows an edge-enhanced Air Force resolution chart. Fig. 1 (b) shows an edge-enhanced picture of a continuous-tone image. In this picture, the outline of the girl's head is clearly displayed and the halo in the background is reduced to a ring.

Further experiments using this real-time edge-enhancing effect in an optical correlator are underway at JPL. The results will also be presented in the meeting.

The research reported in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, as part of its Innovative Space Technology Center, which is sponsored by the Strategic Defense Initiative Organization/Innovative Science and Technology through an agreement with the National Aeronautics and Space Administration (NASA). The work described was also co-sponsored by NASA OAST and the Physics Division of the Army Research Office.

References

1. D. Casasent, F. Caimi, M. Petron, and A. Khomenko, *Appl. Opt.*, 21, p. 3846 (1982).
2. D. Armitage and J. I. Thackara, *Proc. SPIE*, 613, p. 165 (1986).
3. D. Armitage and J. I. Thackara, *Technical Digest of Topical Meeting on Machine Vision of OSA*, 12, p. 58 (1987).



(a)



(b)

Fig. 1 Edge enhancement experiments using a Hughes Liquid Crystal Light Valve.
(a) Edge-enhanced image of an Air Force resolution chart. (b) Edge-enhanced image of a continuous-tone image.

HIGH-GAMMA SPATIAL LIGHT MODULATOR FOR NONLINEAR OPTICAL PROCESSING

C. Warde and H. Lamela-Rivera

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139

H. K. Liu

Jet Propulsion Laboratory
California Institute of Technology, Pasadena, CA 91109.

Abstract

This paper describes how the standard microchannel spatial light modulator can be (a) operated and (b) modified to generate a real-time, high-gamma readout characteristic. The paper also examines the use of such a device in a pulse-frequency modulation optical processor to achieve a wide range of real-time, nonlinear image and signal processing operations such as histogram measurement, density slicing, density-based artificial stereo and gray level image enhancement.

INTRODUCTION

It is well known that coding and modulation techniques are the key to the realization of a large number of two-dimensional, nonlinear image processing operations. The commonly used techniques include linear modulation of an image using Ronchi rulings¹ or sinusoidal gratings,² and spatial pulse-width modulation by contact screens.³ In nonlinear processing, operations such as logarithmic transformation,⁴⁻⁶ exponentiation,^{5,6} level slicing,^{5,7} multiple isophote generation,^{8,9} quantization,¹⁰ pseudocolor,¹¹ A-D conversion,^{8,12} subtraction,^{3,13} and multiple image storage¹ can be realized. Other coding and modulation techniques include optical pulse-frequency modulation and pulse-azimuth modulation.¹⁴ A one-dimensional optical pulse-frequency modulation technique, for example, has been used in conjunction with the variable-grating-mode liquid crystal light valve¹⁵ to demonstrate real-time level slicing.

Unfortunately, the lack of a high-resolution, high-gamma spatial light modulator that would replace photographic film has precluded widespread use of real-time nonlinear optical processing. The microchannel spatial light modulator (MSLM)¹⁷⁻²¹ is a device that is capable of achieving very high gamma. In this paper we describe how the standard MSLM can be (a) operated and (b) modified (Fabry-Perot MSLM) to generate a real-time, high-gamma readout characteristic. The paper also examines the use of such a device in a pulse-frequency modulation optical processor to achieve a wide range of real-time, nonlinear image and signal processing operations such as histogram measurement,²² density slicing,²³ density-based artificial stereo,²⁴ and gray level image enhancement.¹⁷

HIGH-GAMMA CHARACTERISTIC

The desired high-gamma characteristic for nonlinear optical processing is illustrated in Fig. 1. It is a plot of $\log 1/T$ vs $\log E_m$, and it exhibits a sharp threshold exposure, E_{th} . Here T is the modulator readout transmittance and E_m is the modulator input exposure with a threshold of E_{th} . For this negative-gamma characteristic, the readout light remains in the ON state for input exposures

below threshold and is switched off for input levels above threshold. The gamma γ of the system is the slope of the linear region of this curve. That is, γ is given by

$$\gamma = \log(T_1/T_2)/\log(E_2/E_1) \quad (1)$$

Both the standard and the Fabry-Perot MSLMs are capable of high-gamma operation, and both positive and negative gamma characteristics can be achieved. The techniques by which the gamma of an MSLM can be increased include (1) operation of the device in the nonlinear hard-clipped thresholding mode, (2) employing optical feedback around the device, and (3) converting the device to a Fabry-Perot MSLM. The Fabry Perot MSLM inherently generates a high-gamma characteristic when operated in the grid-stabilized saturation mode.

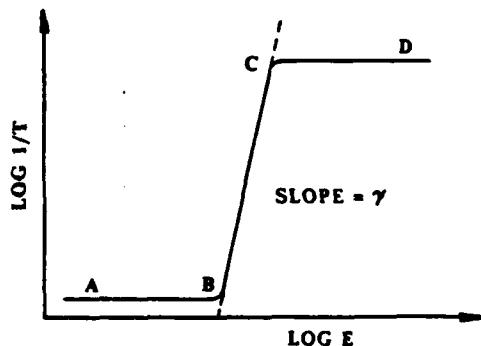


Fig. 1. Desired high-gamma characteristic.

MICROCHANNEL SPATIAL LIGHT MODULATORS

Standard MSLM

The standard Microchannel Spatial Light Modulator (MSLM)¹⁷⁻²¹ is a versatile, real-time optical signal and image processing device that exhibits high optical sensitivity and high framing speed. It consists of a photocathode, a microchannel plate

$$T = I_r/I_i = \sin^2(\Gamma/2) \quad (4)$$

(MCP), a planar acceleration grid and an electro-optic crystal plate, arranged in the proximity focused configuration shown in Fig. 2. The electro-optic plate carries a high resistivity dielectric mirror on one side and a transparent conducting electrode on the other.

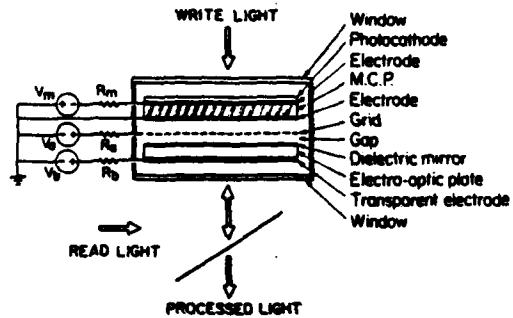


Fig. 2. The optically addressed microchannel spatial light modulator

In the electron deposition write mode, incoherent or coherent write light incident on the photocathode creates an electron image which is amplified by the MCP and proximity focused onto the dielectric mirror surface. The resulting spatially varying electric field in the crystal modulates the refractive index of the crystal. Thus, the readout light which makes a double pass through the crystal is spatially phase or amplitude modulated depending on the crystal cut and readout scheme (polarization or interferometric) employed. The image is erased by appropriately adjusting the device voltages and flooding the photocathode with light so that electrons are removed from the mirror surface by secondary electron emission.

Alternatively, the device can be operated in the reverse mode where a positive-charge image is written by secondary electron emission and erased by adding electrons to the crystal. For Pockels effect crystals, reversing the sign of the image surface charge density leads to contrast reversed images.

In the linear operating mode, the incremental surface charge density $\sigma(E)$ deposited on the crystal is proportional to the exposure E .

$$\sigma(E) = \left[\frac{neG}{hv} \right] E \quad (2)$$

where n is the quantum efficiency of the photocathode, e is the electronic charge, G is the gain of the MCP, h is Planck's constant and v is the frequency of the write light. Additionally, the induced phase change $\Delta\phi$ in the crystal is proportional to $\sigma(E)$ for Pockels effect crystals. The exact expression depends on the type and cut of the crystal employed. For example, in oblique-cut LiNbO_3 ,¹⁸ the electrically-induced phase retardation Γ is of the form

$$\Gamma = \Phi_y - \Phi_x = \frac{\pi \sigma(E)}{\lambda C} [n_e^3 r_y - n_0^3 r_x] \quad (3)$$

where λ is the wavelength of the exposing light, C is the capacitance per unit area of the crystal, n_0 and n_e are the ordinary and extraordinary refractive indices respectively, and r_y and r_x are effective electro-optic coefficients. When such a crystal is read out between crossed polarizers, the transmittance of the crystal is given by

The most recent standard MSLMs have employed LiNbO_3 as the electro-optic crystal. The z-cut crystal has been used for phase modulation and an oblique cut for intensity modulation. Prototype devices employing 50- μm -thick oblique-cut LiNbO_3 crystals have exhibited spatial resolution of approximately 10 cycles/mm at 50% contrast.²¹

Additionally, the internal processing operations achievable with the standard MSLM include multiple-level to two-level intensity image transformation, contrast reversal, contrast enhancement, and binary level image processing operations such as, AND, NAND, OR, NOR, XOR, NXOR. However, when the standard device is operated in the linear mode its gamma is about 2, and this value is too low for nonlinear image processing applications that employ techniques such as pulse frequency modulation. Further details of the principles of operation of the standard MSLM can be found elsewhere.¹⁷⁻²¹

Fabry-Perot MSLM

The Fabry-Perot version of the MSLM employs a crystal that functions as an electro-optically tunable, Fabry-Perot etalon when electrons are deposited or removed from its surface. To fabricate such a crystal standard dielectric mirrors are first deposited on both surfaces of the crystal and then a transparent conducting overcoat is deposited on the readout surface of the crystal. The phase change ϕ which results from the electrically induced refractive index change Δn is, to a good approximation, proportional to the surface charge density $\sigma(E)$ at the crystal surface and to the modulator write light exposure E_m [See Eqs. (2) and (3)].

For a Fabry-Perot etalon with surfaces of reflectivity R , it is well known that the ratio, T , of the total reflected intensity to the incident intensity is given by

$$T = I_r/I_i = \frac{4R \sin^2(\frac{\phi}{2})}{(1 - R^2 + 4R \sin^2(\frac{\phi}{2}))} \quad (5)$$

Shown in Fig. 3 is a plot of T vs ϕ as given by Eq. (5). Note that the reflected readout transmittance of the MSLM approaches zero when ϕ takes on integer multiples of 2π radians.

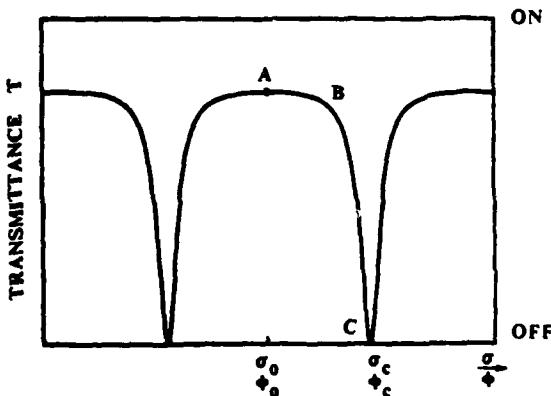


Fig. 3. Interferometric readout intensity characteristic of a Fabry-Perot crystal with $R = 80\%$

Fabry-Perot MSLM With Grid-Controlled Saturation

Because the region A-B-C of the Fabry-Perot characteristic (Fig. 3) closely approximates the corresponding region of the desired high-gamma characteristic (Fig. 1), the Fabry-Perot MSLM will generate the high gamma characteristic when operated in the electron deposition mode with grid stabilized saturation. The device is initially started at the point A (ON state) by precharging the crystal with a uniform surface charge density σ_0 . This is accomplished by flooding the photocathode with light and setting the grid voltage to a value V_0 such that the crystal surface saturates at the corresponding surface charge density σ_0 . Then, by exposing the photocathode with the grid voltage set at V_c , the readout light follows the characteristic in Fig. 3 until the crystal surface voltage reaches V_c with a corresponding charge density of σ_c . Additional exposure has no effect on the readout light because of grid-limited surface charge accumulation. Thus, the high-gamma characteristic of Fig. 1 is achieved.

Hard-Clipped Thresholding Mode

Both the standard and the Fabry-Perot MSLMs can be operated in the hard-clipped thresholding mode to achieve high gamma operation. To operate an MSLM in the hard-clipped thresholding mode with the negative-gamma characteristic shown in Fig. 1, the device is initially biased in the ON state at a peak on its normal ($\sin^2 \Gamma/2$) readout characteristic. This corresponds to the point A on Fig. 2. Then, with the optical input image incident on the photocathode, V_b is ramped downwards at a pre-selected rate V_b that establishes the desired threshold exposure E_{th} . All exposures below E_{th} will be barely recorded because, at their locations, electrons cannot be removed fast enough from the crystal to prevent the gap field from eventually repelling all future primary electrons. Thus, all intensities below threshold will remain in the ON state.

Because the slope of the resulting characteristic is very steep between threshold and saturation, virtually all write light intensities above threshold are written to the maximum charge density which can be set by the terminal value of V_b to coincide with the point C (OFF state) on the device readout characteristic. Thus, operation in the nonlinear hard-clipped thresholding mode leads to the high-gamma characteristic of Fig. 1. Values of gamma in excess of 10 can be easily achieved with both the standard and Fabry-Perot devices using this technique.

Optical Feedback Configuration

Both the standard and the Fabry-Perot MSLM can be operated in the optical feedback configuration shown in Fig. 4 to realize high gamma operation. The device is first biased at the point A on its readout characteristic. As before, this is accomplished by operating the device in the open loop mode (shutter S closed), flooding the photocathode with light and adjusting the grid voltage such that the crystal charges to the uniform surface charge density σ_0 . The image to be processed is then placed in the input plane P_0 and transferred by lens L_0 onto the crystal. The resulting modulated readout signal is then reimaged via the feedback path onto the photocathode of the MSLM. The processed output

image is detected in the plane P_1 .

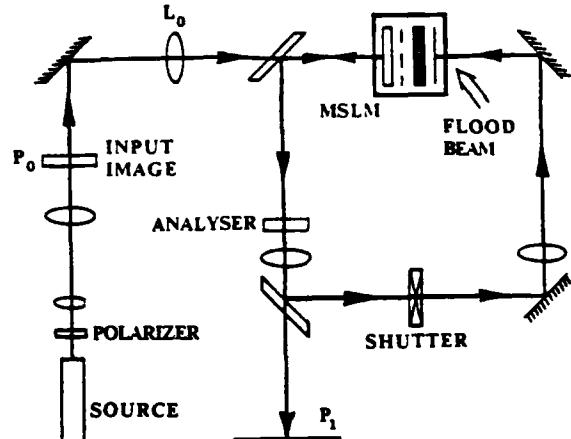


Fig. 4. Optical feedback configuration for high-gamma operation of an MSLM.

In closed-loop operation, for a given fixed uniform exposure E_0 incident on the input image transparency, the high transmittance regions of the input image will at first result in strong feedback to the photocathode. This will quickly result in the deposition of a surface charge density σ_π on the crystal so as to drive the transmittance of the modulator at these points to zero (stable equilibrium). On the other hand, at the low transmittance regions of the input image, the feedback will be weak. This will result in low values of E_m that will deposit only very small amounts of additional charge on the crystal. Since these low exposures will fall below E_{th} , the transmittance of the modulator will remain high. Because the MSLM stores these charge patterns, this feedback configuration will exhibit the high-gamma characteristic shown in Fig. 1 when read out with a separate uniform source. Thus, an input image is transformed into a halftone image by this feedback configuration.

Preliminary Results

A vacuum-demountable Fabry-Perot MSLM was constructed using a 300- μm -thick LiNbO_3 crystal with mirror reflectivities of approximately 50% and 80% on the insulating and conducting surfaces, respectively. The insulating mirror consisted of a multiple layer dielectric coating of SiO_2 and ZrO_2 . The measured readout open-loop, grid-controlled saturation characteristic of the device is illustrated in Fig. 5.

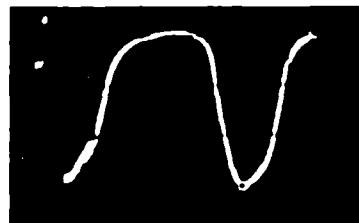


Fig. 5. Measured open loop Fabry-Perot readout characteristic of a vacuum-demountable prototype MSLM with crystal surface reflectivities of 50% and 80%.

Note the improvement over the typical sinc squared characteristic of the standard MSLM. Because of the unmatched reflectivities, the contrast of the measured readout characteristic is not optimum. Nevertheless, the measured open loop gamma of this prototype Fabry-Perot MSLM is 6.2 as compared with a gamma of 2 for the standard MSLM operated in the linear mode. Closed loop gammas are expected to be much higher.

APPLICATIONS TO NONLINEAR PROCESSING

Pulse-Frequency Modulation Technique

The goal of this technique is to perform intensity to spatial frequency conversion on the input image so that subsequent Fourier-plane spatial filtering results in the desired nonlinear operation. The coding or modulation process is generally performed through the use of a contact screen. Contact screens consist of an array of cells with one- or two-dimensional spatial-frequency and transmittance variations $T_{sn}(x,y)$ within each cell.²⁵ An example of the transmittance profile of a specific one-dimensional contact screen is shown in Fig. 6.¹¹

Image modulation is accomplished by sandwiching the original image and the contact screen together and imaging the combined transparency onto a high-gamma (infinite to be ideal) recording medium such as the above-described high-gamma MSLMs. The frequency of the contact screen must be sufficiently high so that within a unit cell the transmittance $T_0(x)$ of the input image transparency is constant. Thus upon readout of the high-gamma recording medium a halftone image of the original is generated.

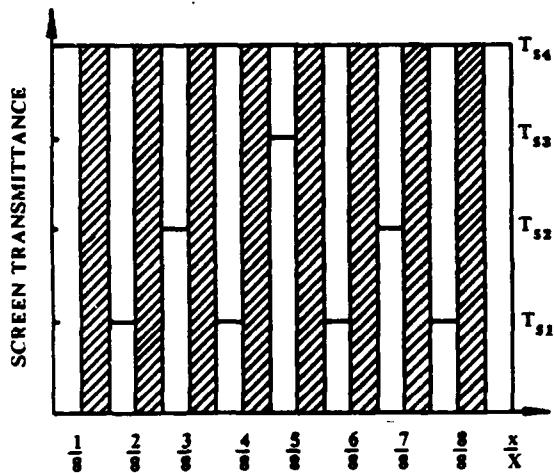


Fig. 6. Transmittance profile of the unit cell of a specific one-dimensional contact screen. Crossed-hatched regions are opaque.

The modulated halftone image is then fed into the input plane of a coherent optical processor, Fig. 7, which separates the various density levels into specific spatial frequency domains within the Fourier plane. With proper design of the screen, spatial filtering in the Fourier plane can be effectively utilized to achieve a wide range of nonlinear image and signal processing operations.

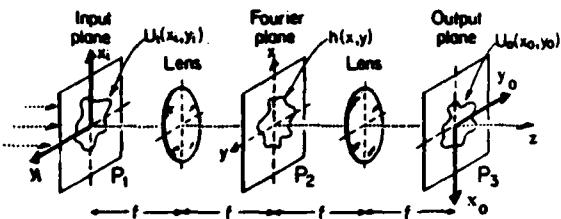


Fig. 7. Standard coherent processing system.

To further illustrate the concept of pulse-frequency modulation, consider the specific case of a one-dimensional contact screen with the unit cell transmittance pattern $T_{sn}(x)$ that is illustrated in Fig. 6. The unit cell of this screen consists of eight slits with four different transmission levels, T_{s1} - T_{s4} . Given a source of constant uniform intensity I_0 , it follows that the exposure incident on the input transparency is given by $E_0 = I_0 t$, where t is the exposure time. Ignoring diffraction losses, the exposure E_m incident on the high-gamma modulator due to the typical slit within the unit cell will be given by

$$E_{mn} = E_0 T_0 T_{sn}; \quad n = 1, 2, 3, 4, \quad (6)$$

where n is a label denoting slits of the same transmittance, and T_0 , the local transmittance of the input image transparency, is assumed to be constant over the unit cell. Note that for the given screen $E_{mn} > E_{m(n+1)}$.

Because the modulator exhibits the threshold characteristic illustrated in Fig. 1, the spatial frequency modulation of its readout light will be dependent on the exposure E_{mn} . For example, given a fixed exposure E_0 incident on the input transparency, if within a given cell T_0 is such that $E_{m4} < E_{th}$, then all E_{mn} for that cell will be below threshold, and the cell will have a uniform transmittance of nearly unity. Thus the modulator readout light from that cell will be spatially unmodulated. That is, its grating carrier spatial frequency would be zero ($f_{gx} = 0$). Likewise, if T_0 is such that $E_{m4} \geq E_{th} > E_{m3}$, then the modulator readout light from that cell will be modulated by a grating with spatial frequency $f_{gx} = 1/X$. For $E_{m3} \geq E_{th} > E_{m2}$, $f_{gx} = 2/X$, ... and for $E_{m1} \geq E_{th}$, $f_{gx} = 8/X$.

When this pulse-frequency modulated image is injected into the conventional coherent optical processor, the diffraction spot at spatial frequency coordinate $f_x = f\lambda/X$ in the Fourier plane will contain the information corresponding to points in the input image with T_0 governed by $T_{s4} T_0 \geq I_{th}/I_0 > T_{s3} T_0$. Here f is the focal length of the Fourier transform lens, λ is the wavelength of the readout source and I_{th} and I_0 are the intensities corresponding to E_{th} and E_0 , respectively. Similarly, the spot at $f_x = f\lambda/2/X$ contains the image information corresponding to $T_{s3} T_0 \geq I_{th}/I_0 > T_{s2} T_0$, etc.

Several experiments can then be performed. For example, by measuring the intensities of the spots in the Fourier plane, a histogram of the original image can be obtained. Second, if different neutral density filters that have higher densities for the spectra with lower intensities are placed behind these spots and all other diffraction orders are stopped, the output image contrast will be enhanced. Third, if a white light readout source is used instead of a monochromatic source, and a set of color

filters are placed at the Fourier plane to selectively pass certain colors, a pseudo-color image will be obtained. Finally, if a set of optical wedges with different wedge angles are placed behind the Fourier plane diffraction spots, artificial stereo pictures will be obtained. These are just a few of the nonlinear processing operations that can be realized with this technique. These experiments are now in progress. Future results will be published elsewhere.

ACKNOWLEDGEMENTS

This research was sponsored in part by the Physics Division of the U.S. Army and the National Aeronautics and Space Administration. The authors would like to thank D. Z. Zhao for helpful discussions.

REFERENCES

1. P. F. Muller, "Linear Multiple Image Storage," *Appl. Opt.* 8, 267 (1969).
2. F. T. S. Yu and M. S. Dymek, "Optical Information Parallel Processing: a Technique," *Appl. Opt.* 20, 1450 (1981).
3. D. Z. Zhao, C. K. Chiang, and H. K. Liu, "Quantitative Optical Image Subtraction by Means of Pulse-width Modulation Through One-dimensional Contact Screens," *Optics Lett.* 6, 490 (1981). Also, "Contact Screen Image Subtraction Technique," *Appl. Opt.* 20, 4234 (1981).
4. H. Kato and J. W. Goodman, "Nonlinear Filtering in Coherent Optical Systems Through Halftone Screen Processes," *Appl. Opt.* 14, 1813 (1975).
5. S. R. Dashiell and A. A. Sawchuk, "Nonlinear Optical Processing: Analysis and Synthesis," *Appl. Opt.* 16, 1009 (1977).
6. M. A. Karim and H. K. Liu, "All Optical Homomorphic Image-Processing System," 7, 371 (1982).
7. S. R. Dashiell and A. A. Sawchuk, "Optical Synthesis of Nonlinear Nonmonotonic Functions," *Opt. Comm.* 15, 66 (1975).
8. T. C. Strand, "Nonmonotonic Nonlinear Image Processing Using Halftone Techniques," *Opt. Comm.* 15, 2394 (1976).
9. H. K. Liu, J. W. Goodman, and J. Chan, "Equidensitometry by Coherent Optical Filtering," *Appl. Opt.* 15, 2394 (1976).
10. S. R. Dashiell and A. A. Sawchuk, "Nonlinear Optical Processing: Nonmonotonic Halftone Cells and Phase Halftones," *Appl. Opt.* 16, 1936 (1977).
11. H. K. Liu and J. W. Goodman, "A New Coherent Optical Pseudo-Color Encoder," *Nouv. Rev. Optique*, 7, 285 (1976).
12. H. K. Liu, "Coherent Optical Analog-to-Digital Conversion Using a Single Halftone Photograph," *Appl. Opt.* 17, 2181 (1978).
13. F. T. S. Yu and A. Tai, "Incoherent Image Addition and Subtraction: A Technique," *Appl. Opt.* 18, 2703 (1979).
14. W. Schneider, "Spatial Pulse Modulation for Coherent Optical Densitometry," *Opt. Acta* 21, 263 (1974).
15. G. Indebetouw, "Production of Color Coded Equidensities Using Non-linear Filtering," *Appl. Opt.* 16, 1951 (1977). Also, "New Method of Pseudo-Color Equidensitometry," *Appl. Opt.* 18, 4206 (1979).
16. B. H. Soffer, et al., "Variable Grating Mode Liquid Crystal Device for Optical Processing," *SPIE* 218, 81 (1980).
17. C. Warde, A. M. Weiss, A. D. Fisher and J. I. Thackara, "Optical Information Processing Characteristics of the Microchannel Spatial Light Modulator," *Applied Optics* 20, 2066-2074, (1981).
18. C. Warde, and J. I. Thackara, "Oblique-cut LiNbO₃ Microchannel Spatial Modulator," *Opt. Lett.* 7, 344-346, (1982).
19. C. Warde, and J. I. Thackara, "Operating Modes of the Microchannel Spatial Light Modulator," *Opt. Eng.* 22, 695-702 (1983).
20. R. Dillon and C. Warde, "X-ray Imaging Characteristics of the Vacuum-Demountable Microchannel Spatial Light Modulator," *Opt. Eng.* 24, 269-273 (1986).
21. T. Hara, Y. Ooi, T. Kato and Y. Suzuki, "Microchannel Spatial Light Modulator with Improved Resolution and Contrast Ratio, *PROC. SPIE*, 613-25, 1986.
22. Ch. Thum, "Histograms of Continuous Tone Image by Transformation and Integration," *Angewandte Optik, Annual Report*, 31 (1980).
23. K. A. Stetson, *Optik* 29, 386 (1969).
24. A. A. Sawchuk, "Artificial Stereo," *Appl. Opt.* 17, 3869 (1978).
25. H. K. Liu, "Halftone Screen with Cell Matrix," U.S. Patent No. 4188225, Feb., (1980).

**Optical-data-processing
properties of a liquid-crystal
television spatial light modulator**

**Hua-Kuang Liu
Jeffrey A. Davis and Roger A. Lilly**

**a reprint from Optics Letters
volume 10, number 12, December, 1985**

Optical-data-processing properties of a liquid-crystal television spatial light modulator

Hua-Kuang Liu

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

Jeffrey A. Davis and Roger A. Lilly

Department of Physics, San Diego State University, San Diego, California 92182

Received July 10, 1985; accepted September 12, 1985

The potential of the extremely inexpensive Radio Shack liquid-crystal television (LCTV) as a two-dimensional spatial light modulator has been investigated. The LCTV modulates the transmission of coherent or incoherent light and can either be electronically addressed through a microcomputer or optically addressed with a TV camera. We have measured the transmission characteristics of the device, examined its diffraction pattern, and tested its use as an input device for an optical correlator. We have discovered that, with proper modifications, it has potential for optical-data-processing applications.

It is well known that optical data processing and optical computing are limited mainly by the capability and availability of real-time devices, such as two-dimensional spatial light modulators (SLM's). For this reason, great efforts by many researchers have been spent in search for usable SLM's. Existing and developing SLM's include the liquid-crystal light valve,¹ the magneto-optic spatial light modulator,² the microchannel spatial light modulator,³ the Pockels readout optical modulator,⁴ and the deformable mirror device.⁵

Many SLM's are either forbiddingly expensive or are still in R&D stages. Recently, several small inexpensive compact liquid-crystal televisions (LCTV's) became commercially available. We have investigated the properties of the Radio Shack LCTV for optical-data-processing applications, and what follows is a brief account of our discoveries.

The LCTV has a 5.4-cm by 4.4-cm screen. The unit consists of a two-dimensional mosaic of raster-scanned liquid-crystal cells, each of which is capable of modulating the coherent or incoherent light transmitted through it. The resolution of the LCTV is determined by the number of cells in the screen. The LCTV under investigation has 146 horizontal elements by 120 vertical elements. Each unit cell is 370 $\mu\text{m} \times 370 \mu\text{m}$ square. The LCTV is equipped with a video input jack, which allows an image to be written electronically with a microcomputer or optically with a TV camera.

For most of our experiments, we have used the computer input mode. When an Apple II+ microcomputer is used, the high-resolution screen with 280 horizontal \times 192 vertical pixels addresses 111 horizontal \times 96 vertical elements of the LCTV. Therefore, there is a 2 to 1 match of vertical elements, permitting exact registration. However, the horizontal registration is off, causing edge blurring.

A TV camera was also used to provide an input

scene, and gray-scale operation was verified by changing the f-stop on the camera.

Each liquid-crystal cell of the TV screen is a 90° twisted nematic liquid crystal sandwiched between parallel polarizers. When no electric field is applied, the plane of polarization for linearly polarized light is rotated through 90° by the twisted liquid-crystal molecules, and no light is transmitted through the second polarizer.

However, under an applied electric field, the twist and the tilt of the liquid-crystal molecules are altered, resulting in a greater fraction of the light retaining the initial polarization direction. As the electric field is increased further, the molecules do not affect the plane of polarization, and all the light passes through the second polarizer. Transmission losses are then limited to the transmission of two polarizer sheets.

Voltage is applied to each pixel through horizontal and vertical line electrodes, which intersect at each pixel. The voltage applied to each pixel is determined by two factors. First, a brightness control determines a bias voltage across every cell that allows the transmission level of the entire screen to be varied. The input signal from the TV receiver, the computer, or the TV camera then varies the individual voltage applied to each cell modulating the individual transmission of each cell, thus creating a SLM with gray-scale capability.

We have performed a series of experiments on the LCTV. Measurements have been made of the optical transmission at the He-Ne laser wavelength as a function of bias voltage across the liquid-crystal unit cells. Control of this voltage is provided by the brightness adjustment on the LCTV. In order to obtain a quantitative measurement, we took the LCTV apart and measured the voltage drop across the variable resistor controlled by the brightness control. (The voltage across a cell is not directly accessible.) The laser

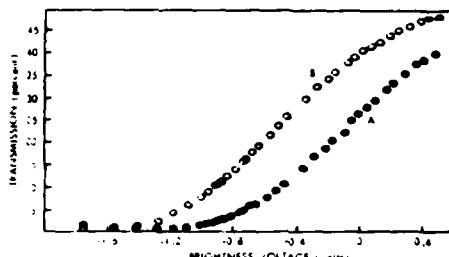


Fig. 1. Transmission measurements of the LCTV as a function of bias voltage. Curve A shows results with the computer pixels turned off; curve B shows results with the computer pixels turned on.

beam was expanded to cover a 1-cm² area of the LCTV screen to average over a large number of pixels. The percentage transmission versus voltage is shown as curve A of Fig. 1. The maximum transmission obtainable is 50%, and the full contrast is about 25/1.

Next, the transmission was measured as the computer wrote uniformly on the LCTV. The computer output places an additional voltage on each cell, modulating its transmission with results shown as curve B of Fig. 1. The computer voltage has the effect of shifting curve A by about -0.4 V. Therefore the operation of the LCTV using a computer input consists of picking a bias point, using the brightness control, and then having the computer change the bias voltage and the resulting transmission at any pixel.

Maximum contrast is obtained by biasing the brightness control at -0.4 V. The high slope of the transmission-versus-voltage curve ensures that the computer will cause a maximum change in transmission. In many SLM applications, it is important to turn all unused unit cells off, and the preferred bias point is at -1.2 V. However, when the pixels are turned on, the transmission increases only to 5%, with the contrast significantly decreased. This change in the voltage applied to each unit cell is set by the TV circuitry. Attempts to increase this voltage swing for higher contrast were unsuccessful. These included varying the video gain and the analog-to-digital converter adjustment inside the TV and externally amplifying the rf signal from the computer. Therefore, without extensive modification, computer operation of the LCTV can provide only minimal binary transmission.

Measurements also have been made of the Fraunhofer diffraction pattern for light incident upon the LCTV; they are shown in Fig. 2(a). As expected from theoretical predictions, the prominent pattern is representative of the grid-matrix type of cell structure of the LCTV elements. Figure 2(b) shows the results when a horizontal grating is written on the LCTV screen with every other line turned on. The expected additional diffraction spots are clearly visible.

Finally, series of measurements have been made using the LCTV as the input screen for a Vander Lugt optical correlator.⁶ The correlation system geometry is shown in Fig. 3. Collimated light from a spatially filtered beam is incident upon plane P₁, where the

input pattern is placed using the LCTV. The Fourier transform plane P₂ is obtained using a 75-cm focal-length lens. The reference beam was also spatially filtered and collimated and then focused using a long-focal-length lens to plane P₃ located beyond plane P₂. This converging reference beam is used to eliminate the final Fourier-transform lens.^{7,8} Matched spatial filters (MSF's) were recorded using a Newport Corporation Model HC-300 thermoplastic recording camera. The focused reference beam synthesizes a Fresnel zone plate within the MSF. Care was taken in recording the MSF to emphasize the spatial frequencies corresponding to the grating information shown in Fig. 2(b) and not the grid information shown in Fig. 2(a). The MSF was made using a computer-generated Ronchi ruling having a period of 10 pixel lines on the high-resolution computer screen.

Plane P₃ contains the correlation between the input pattern and the pattern used in the recording of the MSF. When different computer-generated Ronchi ruling patterns appeared on the LCTV, the correlation spikes were measured using a diode array; results are shown in Fig. 4. Figure 4(a) shows the output when the input Ronchi ruling pattern is identical to that used in making the MSF. The correlation spike is clearly seen. Figure 4(b) shows the results when a computer-generated Ronchi ruling pattern having a period of 24 lines was fed to the LCTV. The correlation spike is reduced, as expected.

The size of the correlation spike decreased as the period of the input ruling varied about the period used

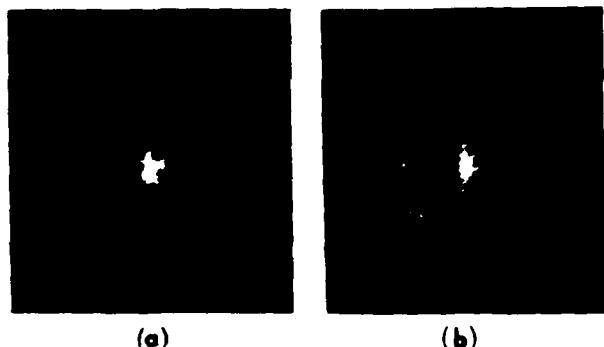


Fig. 2. Fraunhofer diffraction patterns from the LCTV: (a) without (b) with a horizontal grating written onto the screen.

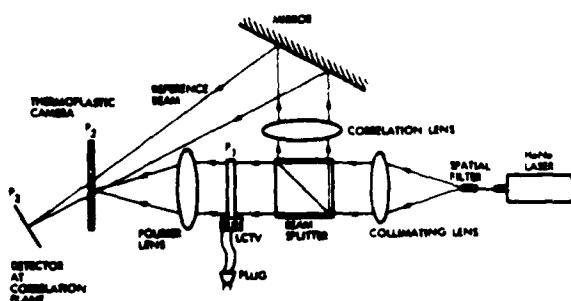


Fig. 3. Schematic of a laser MSF recording and detection system.

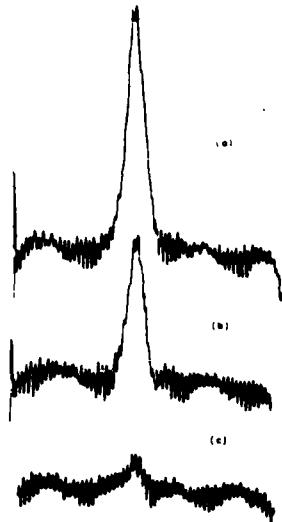


Fig. 4. Cross-correlation signals due to a MSF of a 10-line ruling pattern on the LCTV screen addressed by (a) 10 lines, (b) 24 lines, (c) a blank screen. The vertical axes are of the same scale with arbitrary units. The horizontal axis has a total width of 3.2 mm ($= 128 \mu\text{m} \times 25 \mu\text{m}$).

in making the MSF. Our success in optimizing the MSF is shown in Fig. 4(c), with the LCTV screen uniformly transparent where the correlation spike is almost gone. If the MSF were optimized for the spatial frequencies corresponding to the two-dimensional grid information rather than to the ruling information, the spike would still have been evident.

In conclusion, our experiments show that the LCTV can be used as a SLM. It allows data to be input easily to the LCTV screen from either a microcomputer or a TV camera. The contrast is good when the brightness control is adjusted to midrange. However, if minimum transmission is desired for various unit cells, then the maximum transmission is limited to about 5%. This transmission characteristic compares well with that of the Litton magneto-optical SLM² and should be improvable with changes in the internal electronic circuits of the LCTV.

At the present stage, only relatively coarse objects can be used as inputs for a real-time correlator because

of limitations in the resolution and space-bandwidth product of the LCTV. Nevertheless, TV frame-rate correlation on an input image at a fixed location can be implemented. No shift invariance was observed. This is due to phase nonuniformity of the LCTV, which was discovered by us working with Psaltis's group at the California Institute of Technology campus and also pointed out by one of the reviewers of the manuscript of this Letter. However, the LCTV correlator, after further improvements and modifications, should be useful to robotic vision and space automation programs at NASA and in industry. It is also worthwhile mentioning that the low price and versatility of the LCTV SLM make it attractive for optical-image-processing pedagogical applications.

We are currently investigating other applications for this LCTV, including bipolar modulation,⁹ optical subtraction,¹⁰ high-gamma nonlinear-optical image processing, and optical division applications.¹¹

The research described in this Letter was sponsored by the Physics Division of the U.S. Army Research Office and the National Aeronautics and Space Administration. The LCTV was first called to the attention of one of us (H. K. Liu) by J. G. Duthie. We thank M. Brownell and P. Dickinson for assisting in some of the measurements and M. Hatay for helping to disassemble the LCTV.

References

1. J. Grinberg, A. Jacobsen, W. Bleha, L. Miller, L. Fraas, D. Boswell, and G. Meyer, Opt. Eng. 14, 217 (1975).
2. W. E. Ross, D. Psaltis, and R. H. Anderson, Opt. Eng. 22, 485 (1983).
3. C. Warde, A. M. Weiss, A. D. Fisher, and J. I. Thackara, Appl. Opt. 20, 2066 (1981).
4. R. Sprague, J. Appl. Phys. 46, 1673 (1975).
5. L. J. Hornbeck, IEEE Trans. Electron. Devices ED-30, 539 (1983).
6. A. Vander Lugt, IEEE Trans. Inf. Theory IT-10, 139 (1964).
7. W. Maloney, Appl. Opt. 10, 2127 (1971).
8. M. Bage and M. Beddoes, Appl. Opt. 15, 2830 (1976).
9. D. Psaltis, E. G. Paek, and S. S. Venkatesh, Opt. Eng. 23, 698 (1984).
10. D. Z. Zhao and H. K. Liu, Opt. Lett. 8, 99 (1983).
11. E. Marom, B. H. Sofer, and U. Efron, Opt. Lett. 10, 43 (1985).

On the Progress of The Liquid Crystal Television
Spatial Light Modulator

Hua-Kuang Liu and Tien-Hsin Chao
California Institute of Technology
Jet Propulsion Laboratory
Pasadena, California 91109

ABSTRACT

The principle of operation of the liquid crystal television (LCTV) spatial light modulator(SLM) and the SLM properties of a new high-resolution research module LCTV will be discussed. A comparison of this module with the Radio Shack and Epson LCTV SLMs will be made.

On the Progress of the Liquid Crystal Television Spatial Light Modulator

Hua-Kuang Liu and Tien-Hsin Chao
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

The pocket-sized liquid crystal televisions recently marketed by several Japanese companies have been welcome gifts among friends and found useful in applications such as the view-finders in video cameras. Many optical processing researchers have also begun to use these devices as a spatial light modulators (SLM's) in a variety of ways. It has been demonstrated that LCTV's can be used as optical input devices, logic elements, computer generated hologram recordings, and 2-D phase image representations. The LCTV may be used for these processing applications because it has several attractive features. For example, it is electronically addressable through a micro-computer, and since it is a TV by design, it is naturally addressable by a TV camera and it can be refreshed at a speed of 30 Hz. By definition, when the LCTV is used as a SLM, it can be controlled remotely by an emitting antenna anywhere in the world. So one can imagine that a joint experiment on optical processing could be performed by researchers from Asia, Europe, and the United States through the TV transmitter and receiver linkage. The speed of 30 frames/sec. is sufficient for the present usage because the inputting of image signals from a computer can hardly exceed this speed.

The purpose of this paper is to give a review of the principle of operation of the LCTV SLM, to present the recent findings of the spatial light modulation properties of a high-resolution research module LCTV, and to make a comparison of this module and the Radio Shack, and the Epson modules.

The basic structure of a twisted nematic cell is shown in Figure 1. The liquid crystal molecules are sandwiched between two layers of transparent electrodes¹. An electric field can be applied between these electrodes. In Figure 1(a), it is shown that when the two polarizers are in parallel, due to the rotation of the polarization of the light through the cell, the output is dark when the addressing voltage is off. The output is bright when the addressing voltage is on. The opposite is true when the polarizers are orthogonal as shown in Figure 1(b). In the TV operation, to gain speed, the rotation is much less than 90 degrees and therefore the contrast is reduced. Recently, thin film transistors(TFT) have been built into the LC cells as shown in Figure 2. The function of these TFT's is to hold the addressing voltage and thus to prevent the loss of contrast due to the voltage relaxation effect². The Epson LCTV has TFT's in the cells.

Ordinary LCTV's that are commercially available have a total number of pixels of not over 20,000 and a pixel size of not less than 300 by 300 square micrometers. The characteristics of these types of LCTV SLM have been reported during the past couple of years in the literature. Recently, we have experimented with a research-type LCTV SLM that is not yet commercially available. This LCTV has a screen dimension of 21.56mm x 28.512mm having 220 x 648 total pixels. However, since it is designed for color TV, the number of effective pixels equals one third of the total to account for the three primary colors. For our test, the color filters in the screen have been removed to achieve black-and-white transmission. Under this condition, the LCTV still has a spatial resolution that is about an order of magnitude higher than the existing models on the market.

After proper correction of the non-uniformity of the aperture, due to the uneven structure of the multiple-layer TV screen, we measured the Fourier transform of the high-resolution LCTV and recorded the result which is shown in Figure 3(a). For comparison, a similar result of that of a previous Radio Shack LCTV SLM is shown in Figure 3(b).

To see the other characteristics of this high-resolution LCTV, we have also measured the angular rotation of the liquid crystal molecules as well as the corresponding light transmission and contrast ratios as a function of the bias voltage of the TV. The contrast versus bias voltage is shown in Figure 4. As it can be seen, when the bias voltage is 14 volts, the maximum contrast ratio is only about 6:1. The poor contrast is probably due to two reasons. One is that in order to maintain the 30 Hz TV frame rate, the maximum angle of rotation of the nematic liquid crystals must be relatively reduced to cover the increased number of pixels. In addition, we have found that the laser beam becomes elliptically polarized after it passes through the liquid crystal with a biased voltage as high as 14 volts. This could cause a reduction of the contrast.

To those who question the value of this device because of its low cost, we would respond as follows: If we counted all the development cost that the display community has invested in this device, the cost of each unit of the LCTV would probably be much higher than any of the other existing SLM's on the market today. An example is the research module under test in our laboratory; it was extremely expensive to develop. Furthermore, a significant improvement can be made on this device with regard to both contrast and speed. An LCTV with contrast ratio of over 100, operating at TV frame rate and of high resolution is being developed in Japan. However, if we hope that the LCTV's will serve as SLM's operating according to optical processing specifications, intensive research work in this area should be encouraged. For example, the following additional specifications are desired: light throughput of more than 25%; operating wavelength of 0.6 to 1.3 micrometers; and an overall flatness of within 0.1 wavelength over the aperture and writing speed of 100 sec⁻¹. These goals seem within reach if reasonable development

efforts can be devoted to them.

To present an overview of the latest progress on the LCTV SLM's, we have summarized the characteristics of the Radio Shack, the Epson, and the high-resolution module LCTV SLM and listed the data in a Table of Comparison. This Table will be presented in the meeting.

The authors would like to thank Shigeru Morokawa for providing JPL with the experimental LCTV modules and for many helpful discussions. The research reported in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, as part of its Innovative Space Technology Center, which is sponsored by the Strategic Defense Initiative Organization/Innovative Science and Technology through an agreement with the National Aeronautics and Space Administration (NASA). The work described was also co-sponsored by NASA OAST and the Physics Division of the Army Research Office.

References

1. W. E. Haas, "Liquid Crystal Display Research: The First Fifteen Years", *Mol. Cryst. Liq. Cryst.*, Gordon and Breach Publishers, Inc. 94, 1(1983).
2. S. Morozumi, "Application to the Pocket Color T.V. - TFT array -", *Electronic Magazine*, Ohmsha, Ltd., Tokyo, Japan, 30, No. 6, 39(1985).

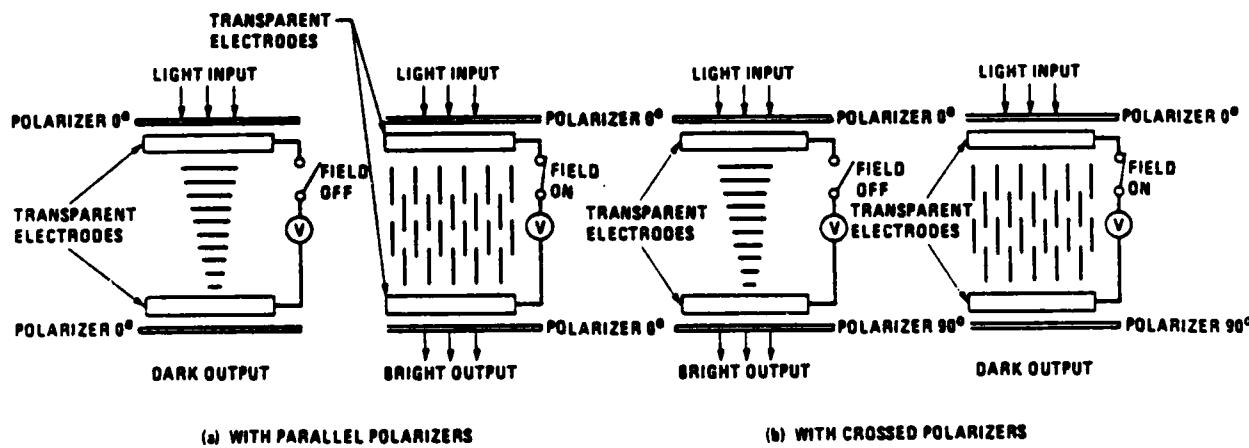


Figure 1. The basic structure and operating conditions of a twisted nematic liquid crystal cell.

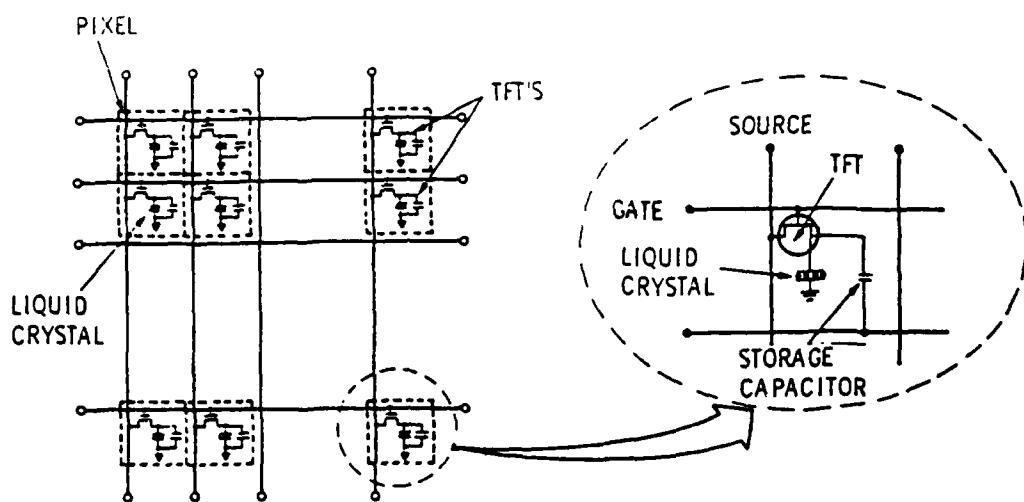


Figure 2. Thin film transistor (TFT) controlled twisted nematic crystal arrays.

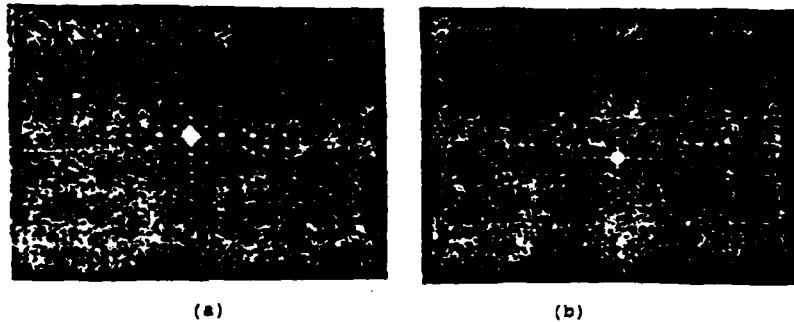


Figure 3. (a) Fourier spectrum of a research-module high resolution LCTV. (b) Fourier spectrum of a Radio Shack LCTV.

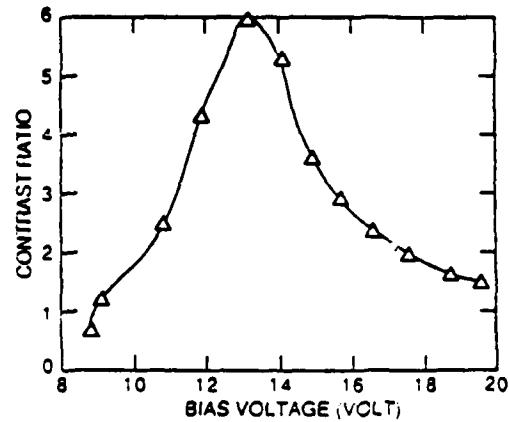


Figure 4. Contrast ratio versus bias voltage of a research-module high-resolution LCTV

**Real-time computer-generated
hologram by means of
liquid-crystal television spatial
light modulator**

**Fai Mok
Joseph Diep
Hua-Kuang Liu
Demetri Psaltis**

**a reprint from Optics Letters
volume 11, number 11, November 1986**

Real-time computer-generated hologram by means of liquid-crystal television spatial light modulator

Fai Mok

Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125

Joseph Diep

Department of Optical Engineering, University of La Verne, La Verne, California 91750

Hua-Kuang Liu

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

Demetri Psaltis

Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125

Received March 17, 1986; accepted August 8, 1986

A novel use of liquid-crystal television (LCTV) is described. It is shown that, if the phase nonuniformity of the LCTV is corrected by a liquid gate, then a simple computer-generated hologram can be written and coherently reconstructed.

It is well known that spatial light modulators (SLM's) are essential for real-time optical information processing. For that reason, much research has been devoted to them, and many types of SLM have been developed.¹

Most of the existing SLM's are either forbiddingly expensive or still in the research and development stages. Recently an extremely inexpensive liquid-crystal television (LCTV) was investigated in relation to its use in optical data processing.²⁻¹⁰ One advantage of the LCTV is that it can be addressed either electronically or optically. However, because of the thickness variations over the aperture of the screen of the devices, its usefulness for coherent optical processing is limited.

We have examined the thickness nonuniformities of the LCTV screen in a Mach-Zehnder interferometer. We have successfully corrected the nonuniformities by submerging the screen in a liquid gate and have demonstrated the feasibility of using this device to write and reconstruct a computer-generated hologram. The results are briefly reported below.

The operating principle of the LCTV has been described.^{2,3} The device has a 5.4 cm by 4.4 cm screen that consists of a two-dimensional mosaic of individually addressable liquid-crystal cells that contain a 90-deg twisted nematic liquid crystal sandwiched between two glass plates. These plates have, respectively, transparent horizontal-line and vertical-line electrodes deposited on their inner surfaces facing the liquid crystal. Two linear polarizer sheets (polarizers) are glued to the outer surfaces of the glass plates. The polarization axes of these polarizers are parallel.

When no electric field is applied, the plane of polarization for linearly polarized light is rotated through 90 deg by the twisted liquid-crystal molecules, and no light is transmitted through the second polarizer. However, with an electric field applied, the twist and tilt of the liquid-crystal molecules are altered, resulting in transmission of a controllable and variable fraction of light. The input signal to the LCTV can be from a TV receiver, a computer, or a TV camera.

A Mach-Zehnder interferometer was set up to analyze the optical quality of the LCTV screen. A 6.4-cm² (1-in.²) region of the LCTV screen was illuminated by a collimated He-Ne laser beam. The results are shown in Fig. 1(a). It can be seen that more than five fringes are produced, indicating thickness nonuniformities of the LCTV screen amounting to about 0.046 μm per pixel. It has been found that the phase nonuniformity is due mainly to the poor quality of the polarizer sheets. Because of the nonuniformity, Fraunhofer diffraction patterns show fuzziness in the various orders.³

We investigated two ways of removing the phase distortions. The first was to correct the phase variation by a phase-conjugation method. We used a 10-mW He-Ne laser to produce a holographic filter of the LCTV screen and addressed it with a phase-conjugated beam projected through a holographic plate. Unfortunately, the beam reconstructed through the LCTV was too weak to be observed visually because of the low diffraction efficiency of the hologram and the low power of the laser. The effectiveness of the phase-conjugation technique was, therefore, not conclusively shown. However, Casasent and Xia have proved that

this is a viable way of correcting the phase-nonuniformity problem.

We then used a liquid-gate correction technique. The LCTV screen was submerged in a liquid gate (NRC Model 550) filled with an index-matching non-conductive mineral oil with a refractive index of about 1.45. In this case, the optical path length through the liquid gate was made nearly constant. Because it is hard to match simultaneously the different indices of the LCTV screen (glass plates plus polarizers), a small amount of phase variation remained. Nevertheless, we have found that most of the thickness variation of the TV screen was corrected. Figure 1(b) shows an interferogram of the liquid gate without the LCTV screen. Note that much less than one fringe appears across the illuminated region when the liquid gate (filled with mineral oil) is placed in the path of the Mach-Zehnder interferometer.

Figure 1(c) is an interferogram of the LCTV screen submerged in the same liquid gate.

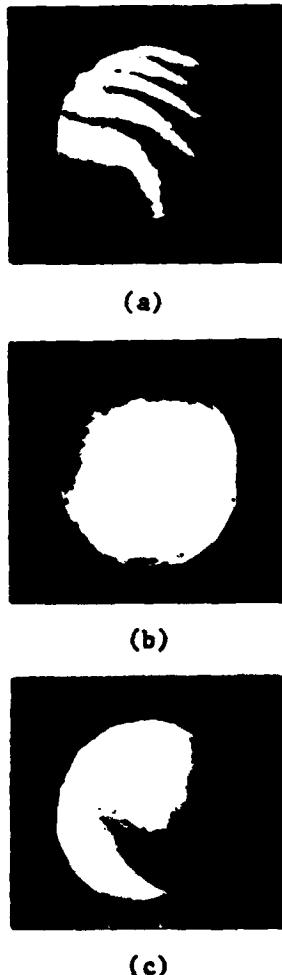
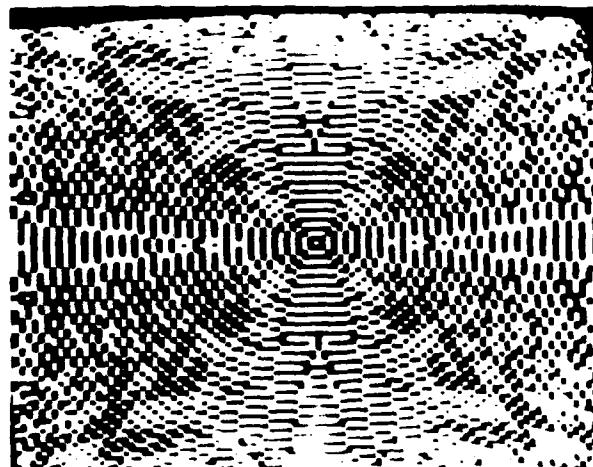
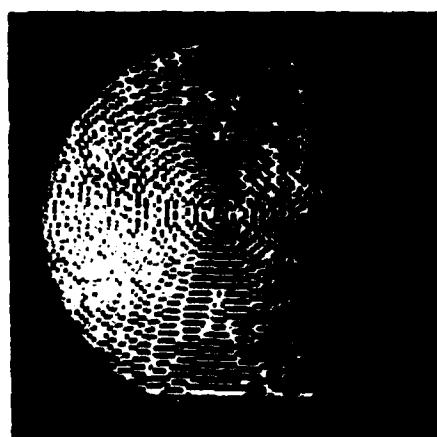


Fig. 1. Phase-nonuniformity correction of the LCTV in a Mach-Zehnder interferometer (MZI) as shown by (a) an interferogram with LCTV in a MZI, (b) an interferogram with liquid-filled liquid gate in a MZI, and (c) an interferogram with LCTV submerged in liquid gate in a MZI.



(a)



(b)



(c)

Fig. 2. Computer-generated hologram of a circle: (a) on a conventional TV monitor, (b) on LCTV; (c) coherent-optical reconstruction.

With the phase distortion of the LCTV mostly corrected, several coherent-optical data-processing experiments can be performed, including an optical matched-filtering experiment.³ We thought it interesting to try to write a computer-generated hologram

on the LCTV and reconstruct it through coherent light. A 100×100 pixel binary phase kinoform¹¹ of a circle was first generated using an IBM PC computer. Figure 2(a) shows the kinoform displayed on a conventional TV monitor. The kinoform was produced by binary encoding of the phase, while the amplitude was kept constant, at each pixel of the Fourier transform of a circle. Basically, this kinoform consisted of on-axis, phase-modulated patterns written into the liquid-crystal layer through the electrodes controlled by the computer output. When the collimated laser light was transmitted through the liquid-crystal layer, the phase of the light was modulated. The output analyzer-polarizer served the function of transforming the phase variations into amplitude variations. However, only the vertical resolution of the kinoform matched that of the LCTV exactly. The LCTV interpolated the missing horizontal points to form an image as shown in Fig. 2(b).

The LCTV, submerged in the oil-filled NRC liquid gate, was then placed in an optical Fourier-transform system. The resulting holographic reconstruction of the circle is shown in Fig. 2(c). As expected, the elliptical appearance of the circle is due to the nonunity aspect ratio of the high-resolution format of the IBM PC. A zero-order dc spot has been blocked for clearer image presentation. The high-intensity dc spot not shown here is believed to be caused by the nonconstant rotations of polarization of the pixels of the same logic level.

To show the necessity of using the liquid gate in the present work, we also recorded a binary phase hologram of the letter O on a piece of film and placed it in front of the LCTV screen with the TV electronics turned off. The reconstruction of this composite filter was fuzzy. After the LCTV was removed, the reconstruction became well defined. Hence we may conclude that the unimproved LCTV can introduce enough phase distortion to cause failure in reconstruction.

The results described in this Letter demonstrate the usefulness of an inexpensive LCTV as a spatial light modulator for coherent-optical processing in the writing and reconstruction of a single computer-generated hologram. Some similar results were achieved by

Psaltis *et al.*, who used a magneto-optics device for reconstruction of a hologram.¹² These findings should make it much easier and less expensive for many optics laboratories to conduct experiments on coherent and incoherent optical processing.

We are currently investigating the use of a large-memory correlator¹³ with the LCTV as a real-time incoherent-to-coherent transducer and with its zero order suppressed.

The research described in this Letter was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Physics Division of the U.S. Army Research Office and the National Aeronautics and Space Administration.

We would like to thank J. W. Goodman of Stanford University and A. Johnston of Jet Propulsion Laboratory for helpful discussion on the use of liquid in the liquid gate.

J. Diep is an academic part-time employee at Jet Propulsion Laboratory, California Institute of Technology.

References

1. G. R. Knight, in *Optical Information Processing*, S. H. Lee, ed., Vol. 48 of Topics in Applied Physics (Springer-Verlag, Berlin, 1981).
2. J. A. McEwan, A. D. Fisher, P. B. Rolsma, and J. N. Lee, *J. Opt. Soc. Am. A* 2(13), P8 (1985).
3. H. K. Liu, J. A. Davis, R. A. Lilly, *Opt. Lett.* 10, 635 (1985).
4. D. A. Gregory, *Appl. Opt.* 25, 467 (1986).
5. M. Young, *Appl. Opt.* 25, 1024 (1986).
6. A. M. Tai, *Appl. Opt.* 25, 1380 (1986).
7. D. Casasent and S. F. Xia, *Opt. Lett.* 11, 398 (1986).
8. H. K. Liu, S. Y. Kung, and J. Davis, *Opt. Eng.* 25, 853 (1986).
9. F. T. S. Yu, J. Jutamulia, and X. L. Huang, *Appl. Opt.* 25, 3324 (1986).
10. J. A. Davis, R. A. Lilly, K. D. Kreng, and H. K. Liu, *Proc. Soc. Photo-Opt. Instrum. Eng.* 613, 245 (1986).
11. L. B. Leesem, P. Hirsch, and J. A. Jordan, Jr., *IBM J. Res. Dev.* 13, 150 (1969).
12. D. Psaltis, E. G. Paek, and S. S. Venkatesh, *Opt. Eng.* 23, 698 (1984).
13. D. Gregory and H. K. Liu, *Appl. Opt.* 23, 4560 (1984).

Optical pseudocolor image enhancement with real-time large screen display

Tien-Hsin Chao
Hua-Kuang Liu, FELLOW SPIE
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

Abstract. A real-time broadband pseudocolor image enhancement technique using a liquid crystal television spatial light modulator (LCTV SLM) is described. Three different schemes to modulate the gray-scale distribution of an input scene by an LCTV SLM through varying the orientations of its analyzer are presented. These schemes are incorporated into a compact white light optical projection system that is able to display, on a large screen, broadband pseudocolor images of real-world input scenes. Experimental results demonstrating the feasibility of the technique are reported.

Subject terms: optical signal processing; industrial applications; real-time image enhancement; pseudocolor encoding; liquid crystal television spatial light modulators; white light processing.

Optical Engineering 27(5), 393-398 (May 1988).

CONTENTS

1. Introduction
2. Gray-scale modulations using an LCTV SLM
3. Optical pseudocolor density encoder
4. Experimental results
5. Conclusions
6. Acknowledgments
7. References

1. INTRODUCTION

Optical pseudocolor density encoding is a processing technique that maps multiple gray levels contained in a black-and-white image into a broad band of colors. Since human observers perceive image variations in colors better than in gray levels (several thousand versus fewer than 20),¹ pseudocolor encoding techniques are useful for subtle gray-level discrimination. This technique can be used for NASA space exploration mapping analysis, earth resources surveys, military intelligence studies, medical diagnosis, and industrial inspection applications.

Pseudocolor encoding of a black-and-white image can be achieved by halftone screen encoding. A halftone encoded image transparency is illuminated by two superimposed laser beams of different wavelengths² or by a collimated white light beam.³ By selective spatial filtering of the different color components contained in the various diffraction orders in the Fourier plane, a pseudocolor density encoded image can be obtained in the output plane. Pseudocolor encoding can also be achieved with grating encodings.⁴⁻⁶ In this method, a positive (original) image, a negative image, and a product of the two are sequentially contact printed through a grating onto a black-and-white photographic film. The orientation of the grating varies with each exposure. For retrieval of the pseudocolor image, the encoded image transparency is illuminated

with a white light plane wave. After color spatial filtering at the Fourier plane, each of the three encoded image components passes through a primary color filter. Although a broad band of pseudocolor can be obtained satisfactorily in the output plane by these encoding techniques, the fact that they are not real-time techniques has severely limited their applications.

Recently, several real-time pseudocoloring density encoding techniques have been reported. One technique utilizes a contrast-reversal spatial filter to obtain a negative image. This negative image is carried by a laser beam of one wavelength and superimposed with a positive image of a different wavelength. The mixture of these two images generates a color-coded image in real time.⁷ Pseudocolor encoding can also be achieved by a method based on the scattering properties of the photographic film.⁸ A positive, a negative, and a bidirectionally modulated image can be obtained simultaneously by illuminating an original image transparency with oblique transmission and diffusive reflection. Coding each of the three images with a primary color and superposing them produces a pseudocolor image. Pseudocoloring has also been obtained directly by using the birefringent effect of a liquid crystal light valve spatial light modulator (LCLV SLM).⁹ Since the LCLV is biased at a specific frequency and voltage and is read out with a white light beam, the output image is automatically color coded. A similar effect has been reported with an LCTV SLM.⁹

In this paper, we report a real-time optical broadband pseudocolor encoding technique that has great potential for industrial applications. The technique utilizes a pseudocolor encoder consisting of a compact white light projection system and an LCTV SLM. This pseudocolor encoder is able to display broadband pseudocolor images out of real-world input scenes through a TV camera. The output can be easily displayed on a large screen for direct viewing. In Sec. 2, we analyze three types of gray-scale modulation schemes using the LCTV SLM. These three schemes form the basis of the pseudocolor encoding. Then we discuss the optical system design of the pseudocolor encoder. Finally, experimental results of a color-coded gray-scale test pattern and a continuous-tone image are used for demonstration.

Invited Paper SP-108 received Feb. 10, 1988; revised manuscript received March 19, 1988; accepted for publication March 19, 1988; received by Managing Editor March 28, 1988. This paper was presented at the OSA Annual Meeting, Rochester, N.Y., Oct. 1987.
© 1988 Society of Photo-Optical Instrumentation Engineers.

2. GRAY-SCALE MODULATIONS USING AN LCTV SLM

Since its introduction three years ago,¹⁰ the LCTV has been adopted as an inexpensive electronically addressed SLM.

We show that the polarization properties of the LCTV cells can be innovatively utilized for pseudocolor encoding. The basic principle of the screen cells is first presented, and a description of the modulation schemes follows. The screen of an LCTV consists of a large array of liquid crystal cells. One of the cells of the LCTV is shown in Fig. 1(a). Twisted nematic liquid crystals are sealed between two transparent electrodes and glass plates. The cell is sandwiched between a pair of polarizers having their axes of polarization states oriented in parallel. As shown in Fig. 1(b), when no electric field is applied across the electrodes, the twisted nematic liquid crystals rotate the axis of polarization state of the incoming light beams by 90°. The output analyzer, whose polarization state

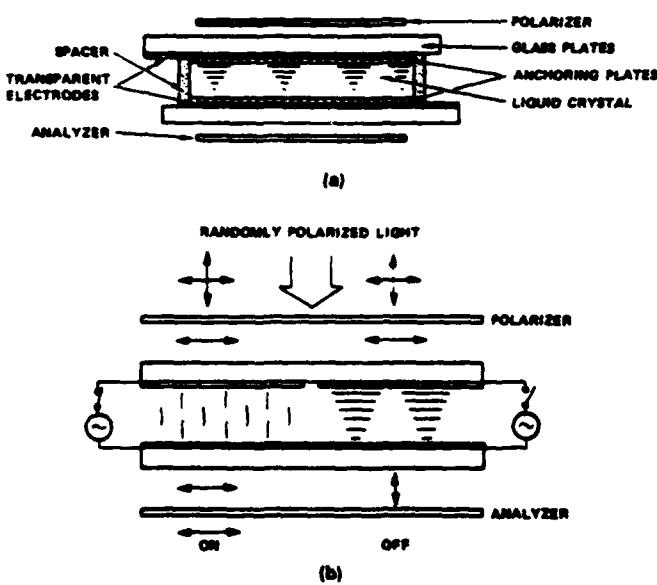


Fig. 1. Spatial light modulation through twisted nematic crystals. (a) Cell structure; (b) two cells with an applied electric field.

is orthogonal, will block the light beams, and the "off" state results. When a sufficiently intense electric field is applied to the electrodes, the orientation of the liquid crystals is twisted back and thus no change is made to the polarization state of the incoming beam. The light beam is able to pass through the analyzer in full amount, and the "on" state results. The gray-scale distribution of its output image can be varied as function of the orientations of its analyzer polarizer. An LCTV is therefore a gray-scale device. Since input scenes are fed into the LCTV through a TV camera or a computer, the array of cells is addressed in a raster scanning mode.

We describe here three different gray-scale modulation schemes. First, as shown in Fig. 2(a), the polarization axis of the linearly polarized input light is rotated by an angle Ω from its off state as it passes through the liquid crystal layer. The rotation angle Ω increases as a function of the increase in the applied electric field intensity, which is proportional to the input image intensity. When the output polarizer is placed perpendicular to the off polarization state, using Malus's law, the light amplitude transmittance t_p of the output image is

$$t_p = K \sin \Omega, \quad 0 < \Omega < \Omega_{\max} \quad (1)$$

where K is a proportionality constant and Ω_{\max} is the maximum value of the angle Ω . Equation (1) shows that, in general, the input is nonlinearly proportional to the output; however, if Ω_{\max} is small, the output light amplitude transmittance is linearly proportional to the input light amplitude transmittance. For convenience, we have plotted the case of small Ω_{\max} in Fig. 2(a). The output image I_{o1} is linearly proportional to the stepped input gray-level pattern I_{in} .

Second, a contrast-reversed image of an input scene can be obtained as shown in Fig. 2(b). In this case, the orientation of the analyzer is perpendicular to the maximum rotatable angle of the liquid crystal (i.e., at $\Omega = \Omega_{\max}$). The light amplitude transmittance of the output image is

$$t_n = K \sin(\Omega_{\max} - \Omega), \quad 0 < \Omega < \Omega_{\max} \quad (2)$$

Equation (2) shows that the output light amplitude distribution is inversely proportional to the input light amplitude distribution. As shown in Fig. 2(b), the output image I_{o2} is a contrast-reversed version of the input image I_{in} .

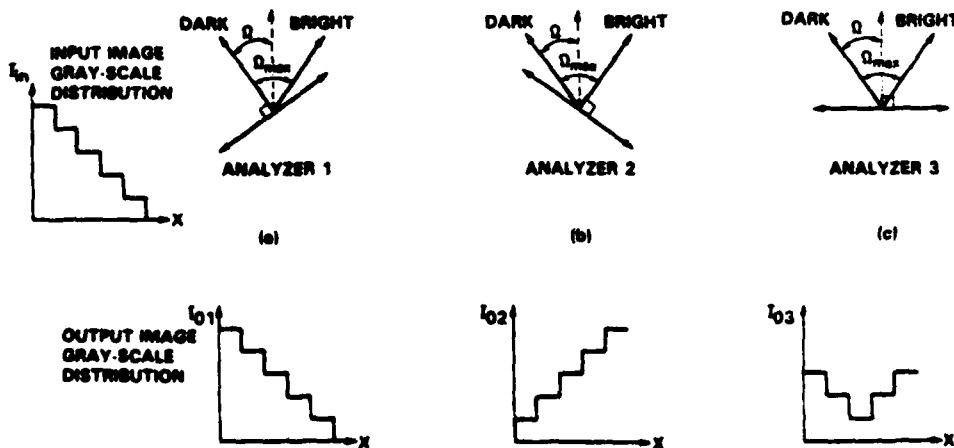


Fig. 2. Gray-scale modulations with respect to analyzer orientations. (a) Positive image; (b) negative image; (c) bidirectionally modulated image.

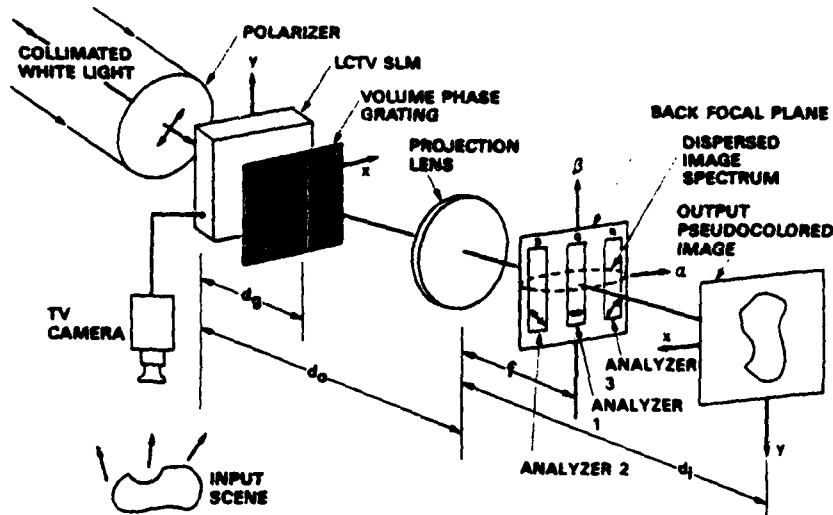


Fig. 3. Diagram of the real-time broadband pseudocolor encoding system.

Third, a bidirectional modulation technique¹¹ is utilized for gray-scale modulation. As shown in Fig. 2(c), when the orientation of the analyzer is perpendicular to the bisector of the two extreme polarization states at $\Omega=0$ and $\Omega=\Omega_{\max}$, the intensity transmittance of the output image becomes

$$t_{bi} = K' \sin\left(\Omega - \frac{\Omega_{\max}}{2}\right), \quad 0 < \Omega < \Omega_{\max}. \quad (3)$$

where K' is a proportionality constant. Equation (3) indicates that by using the bidirectional modulation scheme, we transform the middle point of the input gray scale to the minimum gray level in the output. The highest and lowest levels of the input are transformed to the same highest level in the output. This is illustrated in Fig. 2(c).

It is worth pointing out that in Fig. 2, the linear relationship between the input and output stepped gray-level pattern is true only when the input gray scales fall in the linear region of the transfer characteristic curve of the LCTV SLM. Non-linearity occurs at the toe and shoulder areas of the measured transfer characteristic curves, shown in Fig. 6.

3. OPTICAL PSEUDOCOLOR DENSITY ENCODER

Based on the three schemes of modulating the gray-scale distribution of an input scene through an LCTV SLM as described above, a pseudocolor density encoding method is proposed. By separately encoding each of the positive, negative, and bidirectionally modulated images with one of the primary colors and then superposing them, we can obtain a broadband pseudocolored image. Figure 3 is a schematic diagram of this pseudocolor encoder. A real-world scene is picked up by a TV camera and fed into the LCTV. A collimated white light is used to illuminate the LCTV screen through a polarizer. (As shown in Fig. 3, both the polarizer and analyzer enclosing the LCTV screen are replaced by separately rotatable polarizers.) A phase diffraction grating is placed at a short distance d_g behind the LCTV screen. A projection lens, L , is used to image the input image to the output plane with a magnification. At the back focal plane of

lens L , a set of three analyzers is inserted immediately behind the dispersed image spectrum. Analyzer 1 is placed behind the region of the green color area, analyzer 2 behind the blue, and analyzer 3 behind the red. The remainder of the image spectrum is blocked with an opaque mask. To perform pseudocolor encoding, the orientations of analyzers 1, 2, and 3 with respect to the polarization axis of the impinging spectrum are arranged as depicted in Figs. 2(a) through 2(c). If the input scene amplitude distribution is $t(x,y)$, its corresponding wavelength-dispersed Fourier spectrum $E(\alpha, \beta)$ at the back focal plane behind the analyzers is

$$E(\alpha, \beta) = C \left[T_{pg} \left(\frac{\alpha}{\lambda_g f} \cdot \frac{\beta}{\lambda_g f} \right) * \delta(\beta - \lambda_g fa) + T_{nb} \left(\frac{\alpha}{\lambda_b f} \cdot \frac{\beta}{\lambda_b f} \right) * \delta(\beta - \lambda_b fa) + T_{bir} \left(\frac{\alpha}{\lambda_r f} \cdot \frac{\beta}{\lambda_r f} \right) * \delta(\beta - \lambda_r fa) \right], \quad (4)$$

where T_{pg} , T_{nb} , and T_{bir} represent the positive image Fourier spectrum in green color, the negative in blue color, and the bidirectionally modulated in red color; f is the focal length of the projection lens L ; a is a constant proportional to the grating spatial frequency; C is a complex constant; and the asterisk denotes the convolution operation.

At the output plane, the light intensity distribution may be written as

$$I_o(x,y) = \left\{ K_g \left[\frac{1}{M} t_{pg} \left(\frac{x}{M} \cdot \frac{y}{M} \right) \right]^2 + K_b \left[\frac{1}{M} t_{nb} \left(\frac{x}{M} \cdot \frac{y}{M} \right) \right]^2 + K_r \left[\frac{1}{M} t_{bir} \left(\frac{x}{M} \cdot \frac{y}{M} \right) \right]^2 \right\}, \quad (5)$$

where t_{pg} is the positive image in green, t_{nb} is the negative in blue, and t_{bir} is the bidirectionally modulated in red. K_g , K_b , and K_r are proportionality constants and $M = d_o/d_i$ is the magnification of the projection system. Equation (5) shows

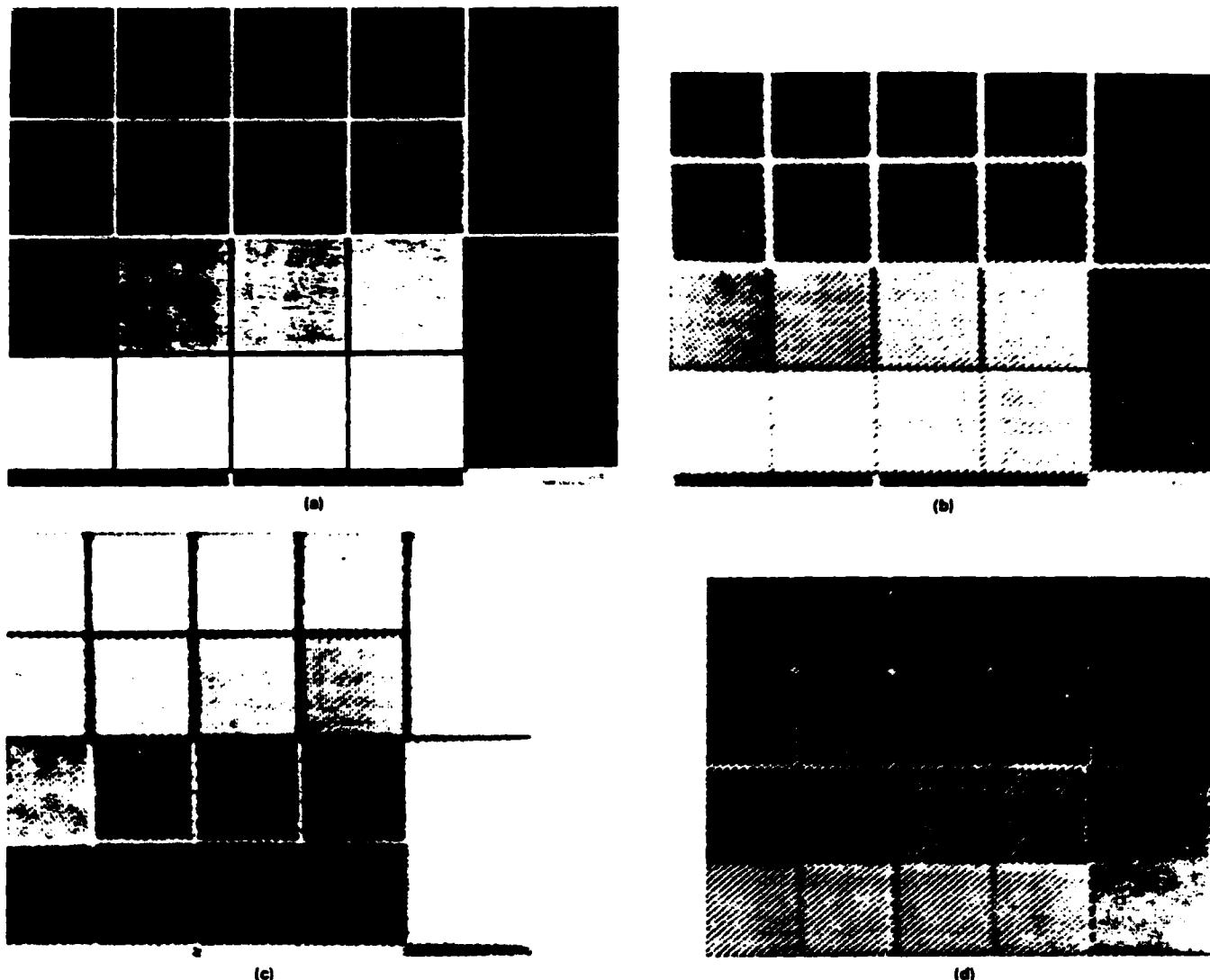


Fig. 4. Experimental demonstration of pseudocoloring (see Fig. 5). Gray-scale test pattern for (a) Original input gray-scale test pattern; (b) positive image of the input pattern obtained from the green spectral band; (c) negative image of the input pattern obtained from the blue spectral band; (d) bidirectionally modulated image of the input pattern obtained from the red spectral band.

that the output image is a superposition of the red, green, and blue of the input image. Each has been gray-scale modulated and thus pseudocoloring is achieved.

4. EXPERIMENTAL RESULTS

We performed some experiments to demonstrate the feasibility of the real-time pseudocolor encoding idea. In our experiments, we used an EPSON LCTV model ET-10 SLM. This LCTV screen contains 220×240 cells. Each contains a thin-film transistor (TFT) that is able to lock up the rotation angle of the liquid crystal after each scanning. Because of the use of the TFT, the resultant contrast ratio is much higher than that of the Radio Shack model.⁹ However, since the EPSON LCTV is a color TV, we addressed it with a black-and-white CCD TV camera to avoid mixing up the pseudocolor with real color. The white light source was a 75 W xenon arc lamp. A dichromated gelatin phase grating was placed about 1 cm behind the LCTV screen to avoid imaging the defects located

on the surface of the grating onto the output plane. The imaging lens was a f/1.8 achromat. Three steps were taken in the preparation of the polarizer filters for pseudocolor encoding. First, a polarizer was placed at the Fourier plane over the green spectral band, with its polarization state oriented parallel to that of the input polarizer. A positive green output image similar to that of the input object was obtained at the output plane. A second polarizer was then placed over the area of the blue spectral band, with its polarization state rotated about 50° away from the first polarizer. This polarizer generated a contrast-reversed blue image at the output plane. Finally, a third polarizer was placed over the red spectral band, with its polarization state oriented between those of the first and the second polarizers (i.e., about 25° away from the first polarizer). The third polarizer produced a red bidirectionally modulated image of the input scene. The simultaneous use of the three encoding polarizers results in a broadband pseudocolored image of a gray-scale input image. With an appro-



Fig. 5. Pseudocolor encoded image of the input pattern obtained by superpositioning the images shown in Figs. 4(b) through 4(d).

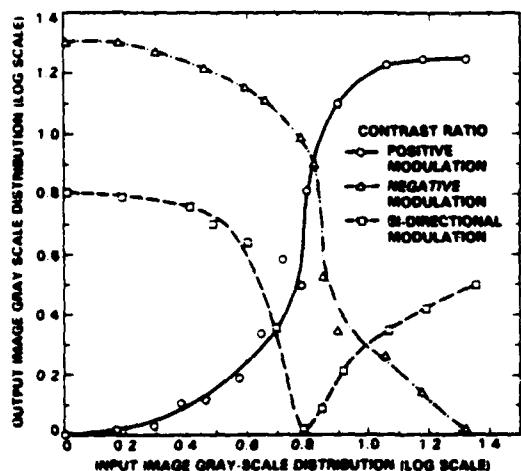


Fig. 6. Output gray-level distribution versus the input gray-level distribution as a function of the orientations of the analyzer. The results were measured from an EPSON model ET-10 LCTV SLM.

priate magnification, the output can be displayed on a large diffuse screen.

In our first experiment, we studied the effectiveness of the three gray-scale modulation schemes by using a photographic gray-scale test pattern as the input object. Figure 4(a) shows an original input gray-scale test pattern; Fig. 4(b) shows a positive image of the test pattern recorded from the green spectral band of the optical encoder; Fig. 4(c) shows a negative image of the input object obtained from the blue spectral band; and Fig. 4(d) shows a bidirectionally modulated image of the input object obtained from the red spectral band. Comparing Fig. 4(d) with Fig. 4(b) we see that the intermediate and the darkest gray scales in Fig. 4(b) were transformed respectively into the darkest and brightest gray scales in Fig. 4(d). The brightest gray scale in Fig. 4(b) remained the same in Fig. 4(d). A color photograph of the resultant pseudo-



Fig. 7. Input image of a golf ball for experimental demonstration of the pseudocoloring of a continuous-tone image (see Fig. 8).



Fig. 8. Pseudocolor encoded image of the continuous-tone input image shown in Fig. 7.

color encoded image made by superimposing the three primary color images is shown in Fig. 5. The brightest pattern was mapped into green color, the darkest pattern was mapped into purple color, and intermediate gray scales were mapped into a broad band of color ranging from orange through yellow to yellow-green. A plot of the measured output image gray-scale distribution versus the input pattern gray-scale distribution is shown in Fig. 6. From Fig. 6 we found that the output image contrast ratios of the positive and of the negative image were both about 20:1. The contrast ratio of the bidirectionally modulated image was only about 8:1.

In a second experiment, we selected a golf ball as the input object. The golf ball was illuminated by a photo lamp from its right rear side. The slowly varying curvature of the ball generated an image containing a wide range of gray scale, as can be seen in Fig. 7. Figure 8 shows the pseudocolor encoded picture

of Fig. 7. Five vivid colors that outlined the contours of five slices of gray levels over the surface of the golf ball can be seen. This result further demonstrates the capability of the pseudocolor encoder in encoding images of real-world input in real time.

5. CONCLUSIONS

We have described three schemes for the modulation of gray-level distribution of an input scene fed into an LCTV SLM and have utilized the principle to build a pseudocolor encoder. We have also described a compact white light projection system that is able to encode and project the pseudocolor image onto a large screen. The principle of pseudocolor encoding is also achievable by using other SLMs such as the LCLV, the SIGHT magneto-optic device (MOD), and micro-channel plate SLM instead of the LCTV SLM. However, from a practical point of view, the LCTV SLM is quite suitable for this application because of its low cost as well as its being electronically addressable. With a large screen display, the pseudocolor encoding system can be used in many industrial applications such as multicolor display, thermal image enhancement, and medical imaging enhancement.

6. ACKNOWLEDGMENTS

This work was performed at JPL, California Institute of Technology, as part of its Innovative Space Technology Center, which is sponsored by the Strategic Defense Initiative Organization/Innovative Science and Technology through an agreement with NASA. Support from the NASA Office of Advanced Science and Technology and the U.S. Army Research Office is also acknowledged.

7. REFERENCES

1. J. A. Mendez and M. Nieto-Vesperinas, "Light scattering by film grain noise: application to grey level optical pseudocoloring," *Appl. Opt.* 22, 2068 (1983).
2. H. K. Liu and J. W. Goodman, "A new coherent optical pseudocolor encoder," *Nouv. Rev. Opt.* 7, 285 (1976).
3. A. Tai, F. T. S. Yu, and H. Chen, "White light pseudocolor density encoder," *Opt. Lett.* 3, 190 (1978).
4. T. H. Chao, S. L. Zhuang, and F. T. S. Yu, "White light pseudocolor density encoding through contrast reversal," *Opt. Lett.* 5, 230 (1980).
5. J. A. Mendez and J. Besco, "Gray level pseudocoloring with three primary colors by a diffraction grating modulation method," *J. Optics* 14, 69 (1983).
6. F. T. S. Yu, X. X. Chen, and T. H. Chao, "White light pseudocolor encoding with three primary colors," *J. Optics* 15, 55 (1984).
7. J. Santamaría, M. Gea, and J. Besco, "Optical pseudocolor density encoding through contrast reversal," *J. Optics* 10, 151 (1979).
8. W. H. Bleha, "Progress in liquid crystal light valves," *Laser Focus*, 1 (Oct. 1983).
9. F. T. S. Yu, S. Jutamulia, T. W. Lin, and X. L. Huang, "R. pseudocolor encoding using a low-cost liquid crystal television," *Opt. Technol.* 19, 45 (1987).
10. H. K. Liu, J. A. Davis, and R. A. Lilly, "Optical-data-processing probe of a liquid-crystal television spatial light modulator," *Opt. Lett.* 1, 1 (1985).
11. D. Psaltis, E. G. Pack, and S. S. Venkatesh, "Optical image correlation with a binary spatial light modulator," *Opt. Eng.* 23(6), 698-704 (1984).



Tien-Hsin Chao received his BS degree from National Chiao-Tung University, Taiwan, his MS degree from Wayne State University, and his Ph.D. degree from the Pennsylvania State University in 1983, all in the field of electrical engineering. He joined the University of Utah as assistant professor of electrical engineering in 1983. Since 1986, he has been with the Jet Propulsion Laboratory, California Institute of Technology. Dr. Chao's research interests include optical information processing, holography, and optical computing. He is an author or coauthor of more than 30 technical papers. Dr. Chao is a member of IEEE and OSA.



Hue-Kuang Liu is a senior research engineer at the Jet Propulsion Laboratory, California Institute of Technology. He is technical group leader of advanced optical processing and holography in the Flight Computer Systems and Technology Section at JPL. Dr. Liu received his BS degree from National Taiwan University, his MS degree from the University of Iowa, and his Ph.D. degree from Johns Hopkins University in 1969, all in the field of electrical engineering. He spent 15 years as professor of electrical engineering at the University of Alabama prior to joining JPL in 1984. Dr. Liu has served as a consultant to NASA, the U.S. Army, and many industrial organizations. His research interests include holography, optical information processing, color and black-and-white printing, and solid-state electronics. Dr. Liu is the author or coauthor of more than 100 technical papers and two book chapters and he holds four U.S. patents and has five U.S. patents pending. He is a Fellow of OSA and SPIE.

An optical inner-product array processor for associative retrieval

S. Y. Kung

Department of Electrical Engineering, University of Southern California
Los Angeles, CA 90089-0272

and

H. K. Liu

Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA 91109

Abstract

In this paper, an inner-product array processor for the associative retrieval problem is presented. First, the algorithm and architecture of the array processor design are discussed. Then an optical implementation scheme is proposed. The matrix model of the associative memory is adopted. In this model, if one of the M vectors is to be reliably recalled, the dimension of the vectors, N , must be much larger than M . By taking advantage of this fact, our result offers a factor of $\sqrt{N/M}$ saving on the matrix elements. More significantly, real-time inputting and updating of the matrix elements can be potentially implemented with existing space-variant holographic elements and recently discovered liquid crystal television spatial light modulators.

I. Introduction

Optical array processors have a great potential because of their parallelism, high computational rates, small size and weight, and their low power dissipation and cost. These processors naturally offer the same strength of VLSI in terms of massive parallelism and yet circumvent the limitation on communication inherent in VLSI processors.

In this paper, we discuss an optical processor design for an associative retrieval processing system. We shall first present a brief history of the evolution of distributed associative memory networks, and the essential algorithmic and architectural design considerations. These then lead to a unique optical implementation of the inner-product array processor for the associative-retrieval problems. There are two memory types in the current computer design; one is location addressable and the other is content addressable. It has been recognized that the content-addressable type is more closely related to how the human brain functions. In fact, a popular approach to its implementation is based on one which resembles that of a neuro-network. Neural signals are trains of pulses with variable frequencies. The task of memory is to reproduce the neural signal at the places where they earlier occurred. This leads to the notion of associative memory network and associative retrieval, whose main features include recognition and error correction based on partial or cluttered input.

II. A Brief Historical Note

Since the 1970's, the correlation matrix model of the distributed associated memory network has been gaining popularity. Notably, Kohonen, Nakano, and Anderson have done independent work in this area in the early 1970's. A brief history is presented below, and key references are provided for researchers who have interests in more details of the evolution of associative memory networks.

Note that the conception of associative memory can trace all the way back to Aristotle's work [370 B.C.] on memory and reminiscence. The cybernetic research work in 1950's on learning digital networks, perceptron, and conditioned connection crossbars [1-3] has paved ways for the modern era of studies on the subject. In the 1970's, the matrix model (as distinguished from the connection model) of memory networks seems to have gained the attentions of many researchers. These work are represented by Morishita [4], Kohonen [5], Nakano [6], Anderson [7]. More recently, Hopfield [8] further extended the previous works to structure a computational model by a notion of energy functions with an outer-product computing model on a iterative basis. To memorize a number of, say M , N -tuple vectors, Hopfield has found through computer simulations that a vector among M can be successfully recalled if M is less than $0.15N$.

III. An inner-product computing model for associative retrieval

One of the most elementary operations implemented by associative memory is the associative recall system. While Psaltis and Farah [9, 10] have presented the optical implementation of associative recall model based on the outer-product of the matrix, we propose a model that is based on the inner product of the matrix with an array of M matrices. The dimensions of each matrix in the array are $\sqrt{N} \times \sqrt{N}$ and the array is $\sqrt{M} \times \sqrt{M}$. We assume that \sqrt{N} and \sqrt{M} are integers. The dimensions of the array are $\sqrt{M} \times \sqrt{N}$ which is less than the dimensions $N \times N$ of the outer-product matrix model since $N > M$.

Referring to Figure 1, the iterative computation is represented by

$$V^* \leftarrow T_\theta [AV]$$

(1)



Figure 1. An iterative computing model for associative retrieval.

where A is a $N \times N$ associative-recalling matrix, which is assumed to be "pre-taught" (i.e., non-adaptive), and $T_\theta[\cdot]$ is a thresholding operation with threshold = θ . In the non-adaptive case, the rank of A is always equal to the total number of information patterns, M . If the columns of A , representing memorized patterns, are nearly orthogonal, then a pattern may be retrieved by its partial information. In order to guarantee a robust and reliable information retrieval, M should be much smaller than the total number of features, N , for example $M < 0.15 \times N$. This implies that A can be expressed as

$$A = \sum_{i=1}^M v_i u_i^T,$$

for a set of vectors (u_i, v_i) ; $i = 1, \dots, M$. (Under the special circumstance of "auto-associative" recall, then $u_i = v_i$.)

The inner product array processor exploits the fact that $M \ll N$, and achieves a considerable amount of hardware savings. The computation in Eq. (1) now becomes

$$A = \sum_{i=1}^M v_i u_i^T,$$

(2)

This can be subsequently decomposed into four steps:

Step (1) $x_i = u_i^T v$, for $i = 1, \dots, M$;

Step (2) $y_i = v_i x_i$ for $i = 1, \dots, M$;

Step (3) $z = \sum_{i=1}^M y_i$;

Step (4) $V^* = T_\theta [z]$;

The four separate steps of this new computing model are depicted in Figure 2.

Referring to the schematic diagram in Figure 2, both the I.P.O. and V.S.O. may be implemented in parallel via an optical array processor. In this case, each N -dimensional I.P.O. (or V.S.O.) may be implemented by one $\sqrt{N} \times \sqrt{N}$ "optical-plane-product," as shown in Figure 3. This means that a much more compact-size optical implementation process is now possible.

IV. An optical implementation of a programmable associative recall system

An optical system for implementing the inner-product associative-retrieval model as

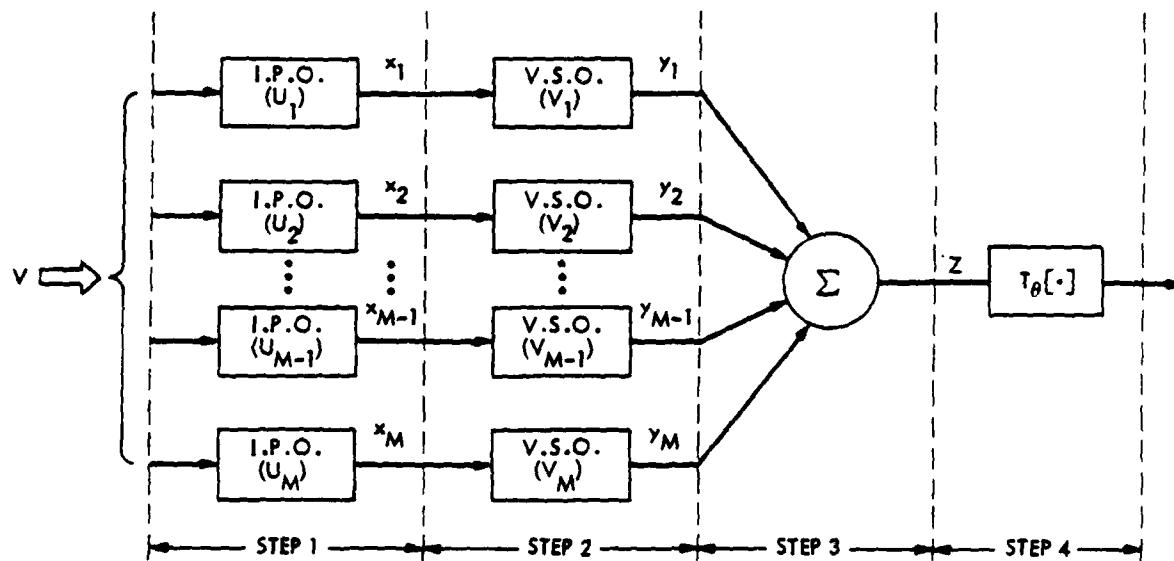


Figure 2. The inner product computing model where I.P.O. represents an inner product operator and V.S.O. represents a vector scaling operation.

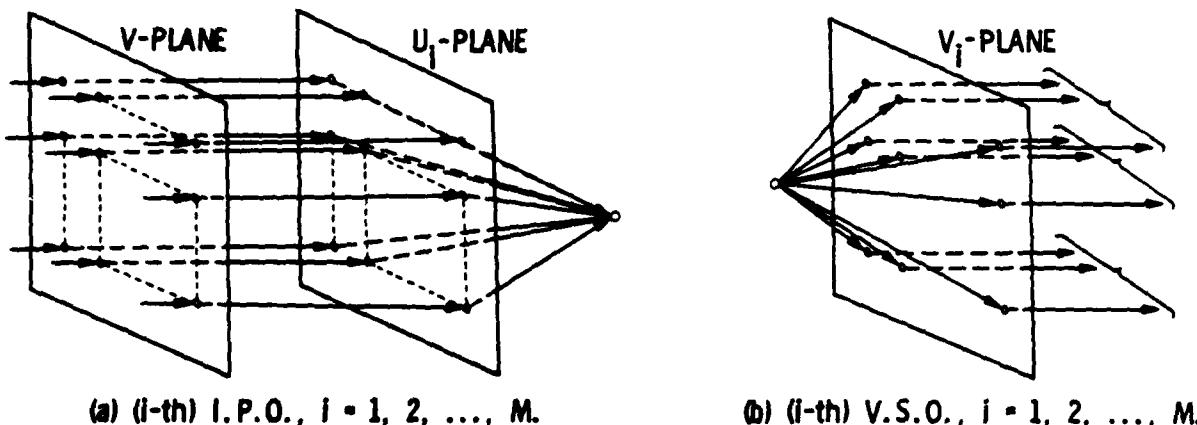


Figure 3. A schematic diagram of the optical implementation of I.P.O. and V.S.O.

discussed above is shown in Figure 4. In the Figure, a liquid crystal television (LCTV) is used as an electronically addressed spatial light modulator (SLM). In fact, other electronically addressable SLM's such as the Hughes CCD-addressed LCLV may also be able to serve the same function as the LCTV. Two multiple-focus holographic lenses¹² (hololens), MHL1 and MHL2 are used as space-variant optical elements for performing the necessary inner-product and summation functions. The function or capability of the MHL's is shown in Figure 5. The holographically-made optical element is capable of replicating an input 2-D image by first focusing it to an array of focal points at its focal plane and then forming an array of images at its image plane. All the replicated images are derived from a single image through the common aperture of the MHL; a capability that cannot be achieved by any currently available refractive elements such as a glass lens.

To set up the optical array processing system, we first display the M known vectors on the LCTV each in a \sqrt{N} by \sqrt{N} matrix format. The nature of the LCTV's transmissive screen is to modulate the input laser light by what is being written according to the matrix-format vector. The $\sqrt{M} \times \sqrt{M}$ matrices being written or displayed on the LCTV screen exhibit its storage capacity and the transmission modulation property of the screen enable these matrices to be multiplied by other matrices. Hence once the writing process is completed, the storage step is done.

To activate the addressing-recalling mode of operation of the system, an input vector can be invoked. The N -tuple vector should be of the same matrix format as that of any one

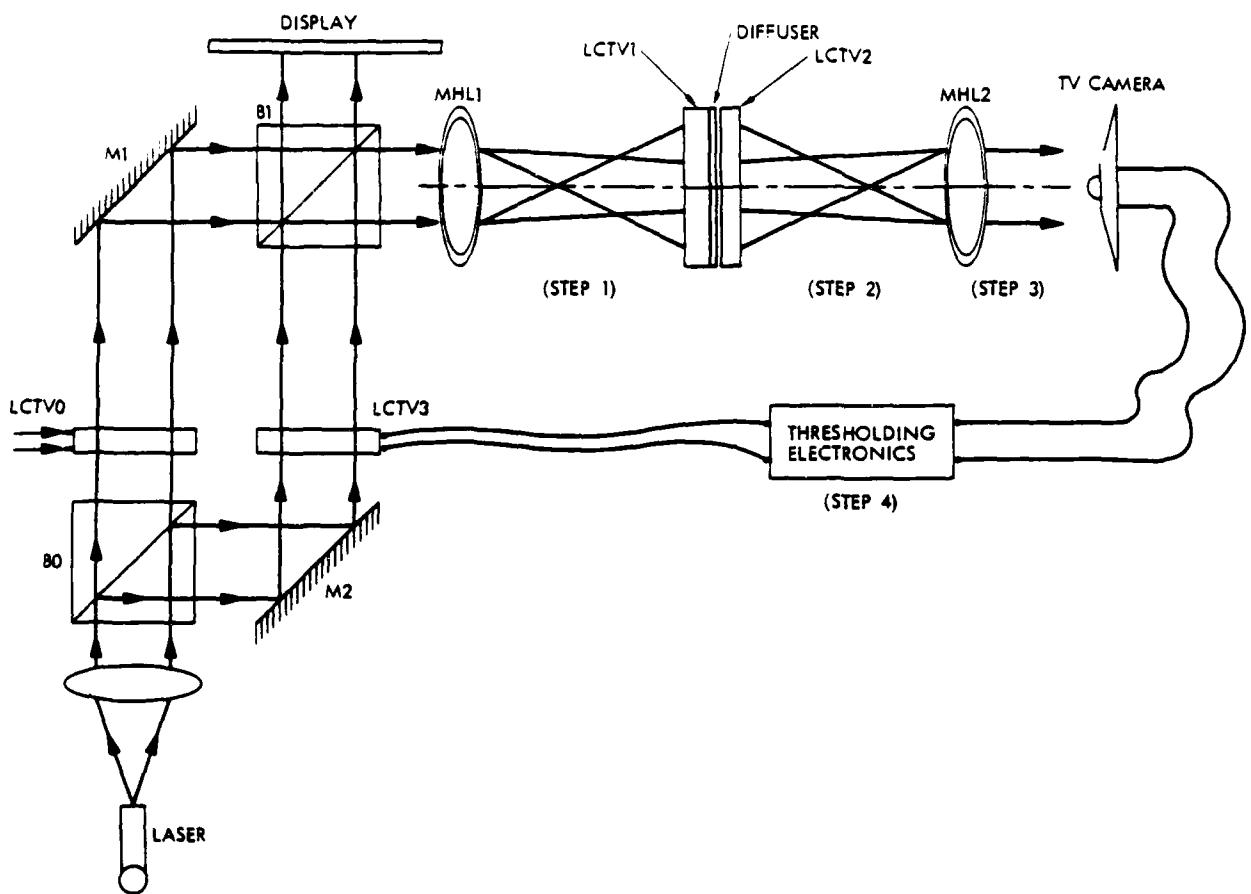


Figure 4. A schematic diagram of an optical implementation of the inner-product array processor.

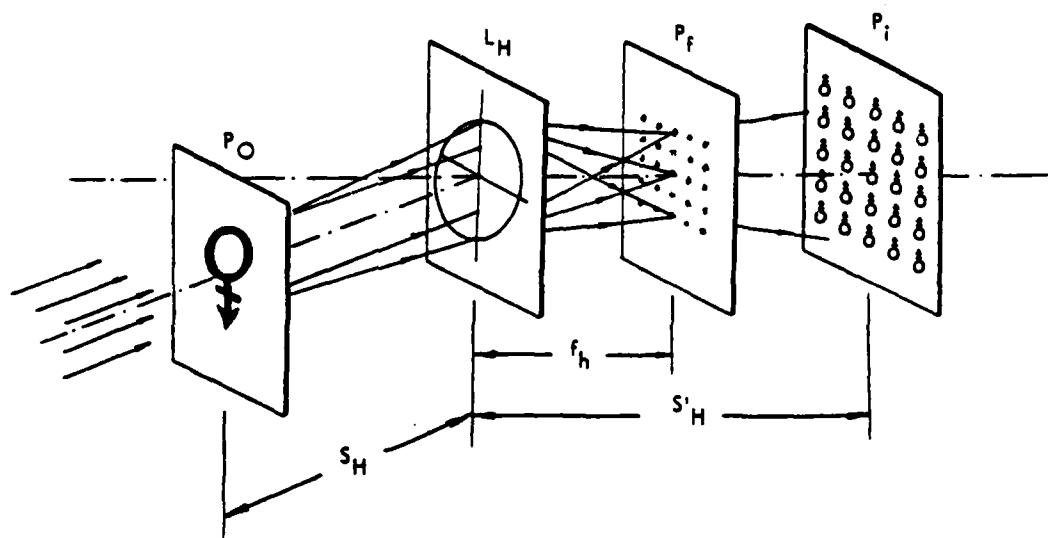


Figure 5. The function of a multiple-focus holographic lens.

can be invoked. The N -tuple vector should be of the same matrix format as that of any one of the M vectors so that it can be written onto LCTV0. The screen of LCTV0 can then be used to modulate the collimated laser beam through the cube beam splitter BO. The information being written on LCTV0 can then be sent to address the M vectors. After the

light carrying the vector information transmitted through B_1 and MH_1 , it is replicated to a \sqrt{M} by \sqrt{M} (here we assume that \sqrt{M} is an integer) array of matrices of $\sqrt{K} \times \sqrt{K}$ dimensions of elements coinciding matrix by matrix with what has been previously stored in $LCTV_1$. The inner-product-type operation is thus partially achieved at this step. To complete the entire inner product process, the multiplied results are followed by a diffuser which averages the light intensity and yield the M inner-product scalars. The scalars are then transmitted through to another array of M known vectors displayed (or written) on $LCTV_2$. (Computationally, this process implies that the M vectors are multiplied with the scalar.) The matrix array on $LCTV_1$ and $LCTV_2$ should be identical in case of auto-associative retrieval. Finally, the weighted vectors (or matrices) are summed up by MH_2 which plays a reversed role as MH_1 . A TV camera or CCD detector array can be used to detect the output. The output should be thresholded by an electronic device. (The thresholded array may also be performed by an optical device if a high-gamma spatial light modulator is available.) If the electronic device is adopted (as shown in Figure 4), the thresholded result can be written on $LCTV_3$ and illuminated by a collimated laser light from B_0 . This thresholded matrix is then used as a feed-back input through R_1 and MH_1 ready for a new iteration process.

The above-mentioned procedure will repeat until a steady-state is reached and exhibited in the display sent also through B_1 . Once the steady-state is reached, the retrieval process is thus completed.

A partial optical set-up of the inner-product associative-retrieval model is shown in Figure 6. In the forefront, a LCTV imbedded in a liquid gate is shown. The electric connection to the LCTV is also visible. At this stage, the set-up is only for the purpose of testing the quality of the LCTV in its applications as a spatial light modulator. The results of the testing are important to the implementation of the associative-retrieval model.

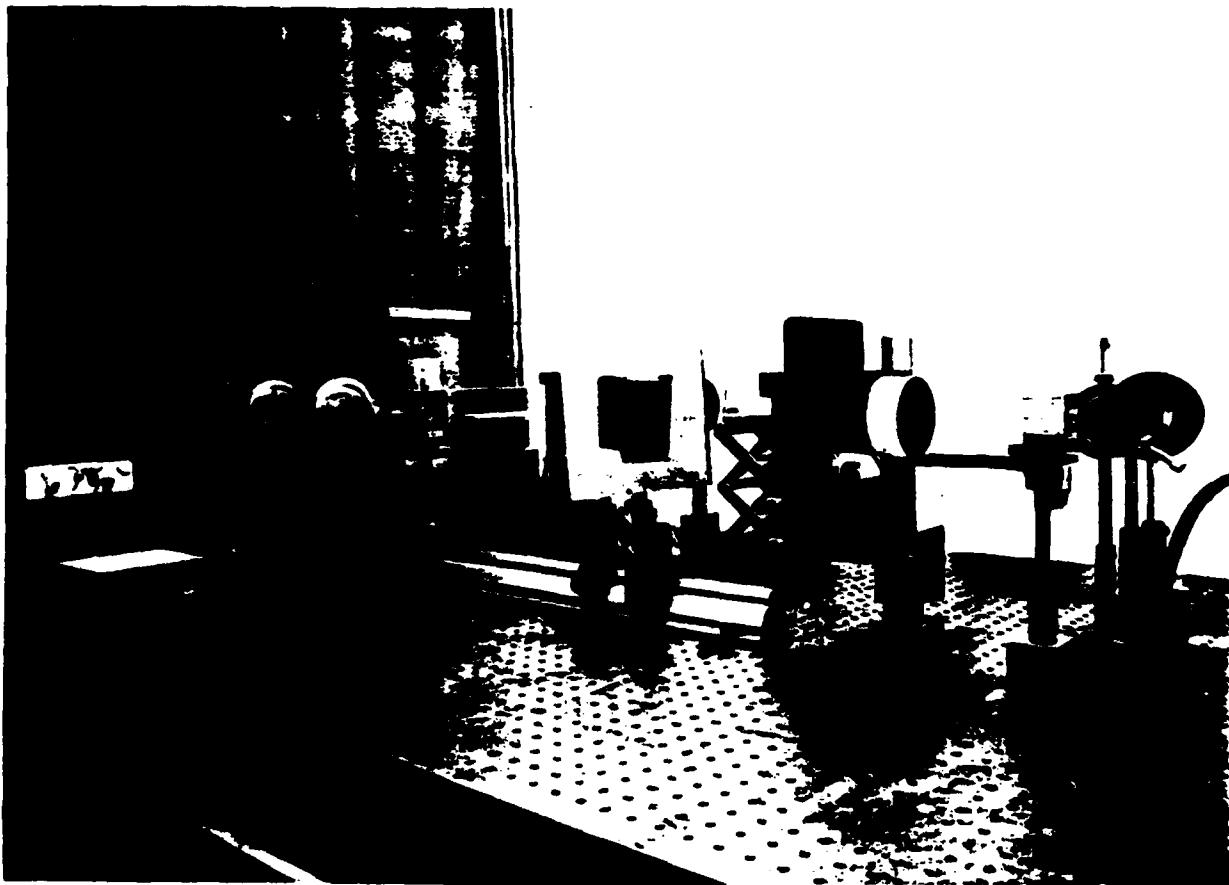


Figure 6. A photograph of a part of the laboratory set-up of the optical inner-product array processor.

V. Conclusions

The principle and implementation of a programmable associative recall system based on the inner-product model have been described. Advantages of the system include:

1. The dimensions of the matrix in the computation have been reduced. For example, in the Kohonen-Nakano-Hopfield model, the dimension of the required memory matrix is $N \times N$. Whereas in the present case we only need an array of M matrices of dimension $\sqrt{N} \times \sqrt{N}$. Numerically, if $N = 100$ and $M = 4$, then a 2×2 array of 10×10 matrices is required in the inner-product model, instead of a 100×100 memory matrix. If $N = 900$, $M = 100$, then a 10×10 array of 30×30 matrices is sufficient to perform the retrieval function. Therefore, one order of magnitude of saving on the memory matrix elements can be achieved and make the optical implementation more feasible.

2. The optical implementation exploits the newly developed inexpensive LCLV as spatial light modulators and multifocus lens as the replicating, holographic interconnecting, and matrix-matrix multiplication processors.

3. Real-time modification of the memory can be achieved by changing the matrix elements through electronically reprogramming the LCTVs.

We are hopeful that the proposed inner-product model, once realized, will be very useful for many associative memory applications.

Acknowledgment

This research was supported in part by NASA, Army Research Office, National Science Foundation Grant ECS-82-13358, and Office of Naval Research Contract No. N000014-85-K-0599.

References

1. B. G. Farley and W. A. Clark, "Simulation of Self-Organization Systems by Digital Computer", IRE Trans. IT-4, 76, (1954).
2. F. Rosenblatt, "The Perception: A probabilistic model for information storage and organization in the brain", Psych. Rev. 65, 386, (1958).
3. K. Steinbuch, "The learning matrix", Kybernetik, 1, 36, (1961).
4. I. Monishita, "Analysis of an adaptive threshold logic unit", IEEE Trans. C-19, 1181 (1970).
5. T. Kohonen, "Correlation matrix memories", Helsinki U of Tech. Report TKK-F-A/30, i.b.i.d. IEEE Trans. C-21, 353 (1972).
6. K. Nakano, "Associatron-A model of associative memory" IEEE Trans. SMC-2, 380, (1972).
7. J. A. Anderson, J. W. Silverstein, S. A. Ritz and R. S. Jones, "Distinctive features, categorical perception, and probability learning: some application of a neural model", Psych. Rev. 84, 413, (1977).
8. J. J. Hopfield, "Neural network and physical systems with emergent collective computational abilities", Proc. Natl. Acad. Sci. USA 79, 2554, (1982).
9. D. Psaltis and N. Farhat, JCO-13, August, 1984, Saporro, Japan.
10. D. Psaltis and N. Farhat, Optics Letters, 10, 98, (1985).
11. H. K. Liu, J. A. Davis, and R. A. Lilly, Optics Letters, 10, 635 (1985).
12. Y. Z. Liang, D. Zhao, and H. K. Liu, Appl. Opt. 22, 3451, (1983).

Real-time optical associative retrieval technique

Hua-Kuang Liu, MEMBER SPIE
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

S. Y. Kung
University of Southern California
Department of Electrical Engineering
Los Angeles, California 90089

Jeffrey A. Davis, MEMBER SPIE
San Diego State University
Department of Physics
San Diego, California 92182

Abstract. A real-time optical associative retrieval technique is presented. The associative retrieval model enables a large amount of data to be stored and recalled by partial information optically in real time. The real-time capability is achieved by using an electronically addressed spatial light modulator based on the pocket-size liquid crystal display television. The potential application of the technique to the perceptive vision requirements in telerobotics for achieving NASA's goals of automation in space is described.

Subject terms: Space Station optics; optical processing; automation and robotics; liquid crystal television spatial light modulator (LCTV SLM); multiple-focus holographic lens (hololens); associative retrieval; real-time robotic vision; correlator.

Optical Engineering 25(7), 853-856 (July 1986).

CONTENTS

1. Introduction
2. Experimental evaluation of the LCTV SLM
3. Optical implementation of a programmable associative recall system using LCTV SLMs
4. Conclusions—relevance to the automation and robotics (A&R) program at NASA-JPL
5. Acknowledgments
6. References

1. INTRODUCTION

NASA-JPL has a major program in automation and robotics (A&R) for planetary exploration. To date, the developments of sensing and perception of the A&R program have been based mainly on digital computer methods, which have the advantage of flexibility and reprogrammability. However, digital techniques are restricted by their limited space-bandwidth product (SBP), especially in the perception of multiple objects with complex structures moving with six degrees of freedom. The inherent parallelism of optical processing and its large SBP therefore naturally offer an alternative and/or complementary approach to digital optical processing for A&R.

In this paper, we discuss an optical system utilizing the pocket-size liquid crystal television (LCTV) as a spatial light modulator (SLM). This system can be used for real-time associative memory and pattern recognition in space applica-

tions. First, we discuss the characteristics of the LCTV. Then we show the potential usage of the LCTV in an inner product associative memory model. Finally, we outline the immediate and long-range goals of work to apply these techniques in the A&R program.

2. EXPERIMENTAL EVALUATION OF THE LCTV SLM

Optical pattern recognition has been limited by the unavailability of real-time two-dimensional spatial light modulators. Among the commercially available SLMs are the liquid crystal light valve (LCLV),¹ the magneto-optic spatial light modulator,² and the microchannel spatial light modulator.³ These SLMs are very expensive. Recently, however, several versions of a small, inexpensive, compact liquid crystal television have become commercially available, and their properties for optical data processing have been explored.^{4,5} Salient features of these devices for use in an associative memory system will first be discussed.

Although various versions of the LCTV are commercially available, most of our experiments have been done with the Radio Shack LCTV (Realistic Pocketvision Cat. No. 16-151 or 16-153) with a 5.4 cm by 4.4 cm screen. The screen has a two-dimensional grid of raster-scanned liquid crystal cells, each of which modulates the light transmitted through it. The resolution is typically 146 horizontal elements by 120 vertical elements that are each 370 μm by 370 μm square. The LCTV comes equipped with a video input jack that allows an image to be written electronically with a microcomputer or with a TV camera.

Each cell of the LCTV screen is a 90° twisted nematic liquid crystal between parallel polarizers. When no electric field is

Invited Paper SS-102 received Feb. 3, 1986, accepted for publication Feb. 3, 1986, received by Managing Editor March 24, 1986
© 1986 Society of Photo-Optical Instrumentation Engineers

applied, the plane of polarization for linearly polarized light incident on the cell is rotated through 90° by the twisted liquid crystal molecules, and no light is transmitted through the second polarizer. However, under an applied electric field, the twist and tilt of the liquid crystal molecules are altered, affecting the rotation angle of the transmitted light and therefore varying the amount of light transmitted through the second polarizer.

Voltage is applied to each pixel through horizontal and vertical line electrodes that intersect at each pixel. The voltage applied to each pixel is determined by two factors. First, an adjustable brightness control applies a uniform voltage to every pixel, allowing the transmission level of the entire screen to be uniformly varied. In addition, the input signal from the TV receiver, computer, or TV camera then varies the individual voltage applied to each pixel, thus modulating the individual transmission of each pixel and creating the picture.

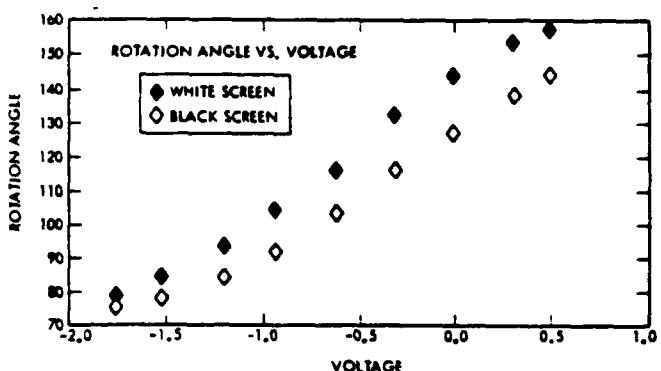


Fig. 1. Rotation angle of light transmitted by LCTV cells versus brightness control voltage. Curves show data for screen when computer addresses each pixel (closed triangles) and when computer turns each pixel off (open triangles).

Experiments studying the polarization of the light transmitted by the liquid crystal cells were performed after the second polarizer was removed from the LCTV. The transmitted polarization state was analyzed with a second rotatable polarizer as a function of bias voltage across the LCTV at the He-Ne laser wavelength of 632.8 nm. The brightness adjustment on the LCTV provided control of this voltage and was monitored using the LCTV control electronics. The plane of polarization transmitted through the liquid crystal elements was essentially linear and rotated as the brightness control voltage was varied, as shown in Fig. 1. The two curves show results when the computer monitor was turning every pixel on and when every pixel was turned off. The two curves essentially differ by a rotation angle of 11° . The transmitted light intensity will then vary upon (1) applying a voltage from either the television camera or computer; (2) changing the brightness control voltage; or (3) varying the angle of the second polarizer. Note that this latter degree of freedom can be changed only if the second polarizer is removed from the LCTV.

Although only binary operation is available under computer control, limited gray scale operation can be obtained by addressing the LCTV using a television camera.

Because the LCTV is not optically flat, there is a phase nonuniformity across the surface that destroys much of its capability as a SLM. This effect was examined by studying the Fraunhofer diffraction pattern produced by the device. Light from a He-Ne laser was collimated and sent through a 36.8 cm focal length lens. A diode-array detector placed in the focal plane of the lens was used to record the Fraunhofer diffraction pattern. Figure 2(a) shows the focused spot produced by the system. Figure 2(b) shows the broadening of the focused spot when the LCTV is introduced into the system. This broadening is evidence of phase nonuniformity across the plane of the LCTV.

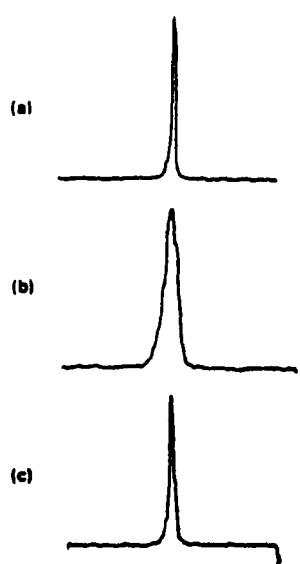


Fig. 2. Fraunhofer diffraction pattern of collimated beam using 36.8 cm focal length lens and diode array detector. (a) Nothing placed in optical path. (b) LCTV placed in optical path. (c) LCTV and liquid gate placed in optical path. Note full horizontal scale is 1.6 mm.

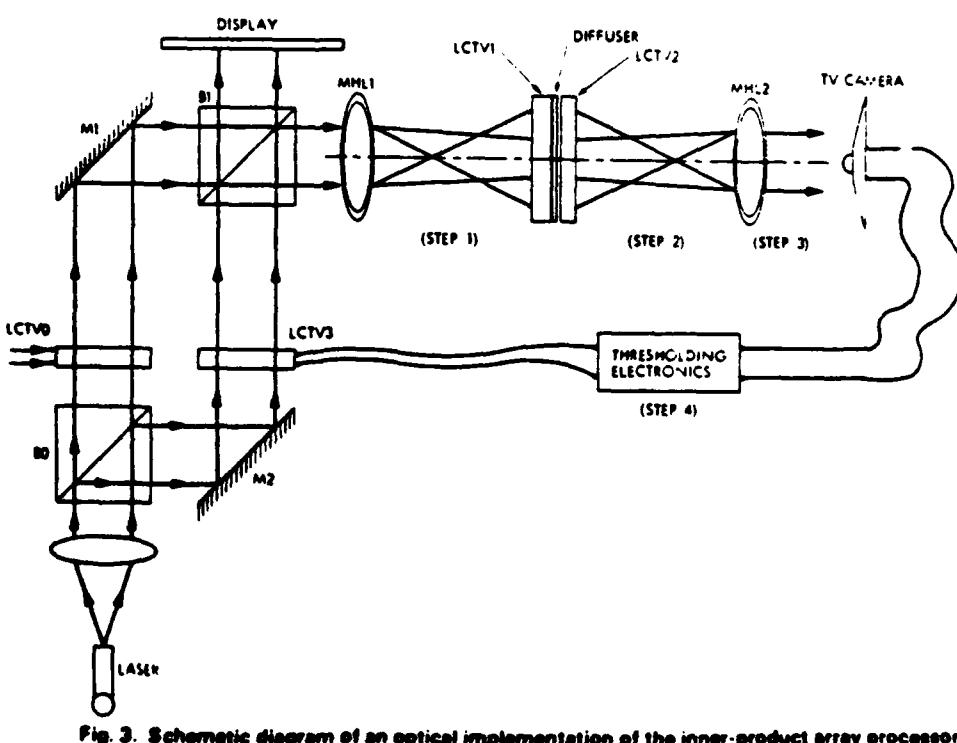


Fig. 3. Schematic diagram of an optical implementation of the inner-product array processor.

These phase variations were corrected by placing the LCTV inside a Newport Corporation Model 550-G liquid gate filled with mineral oil. This step successfully eliminated the phase variations, as shown by the sharply focused spot in Fig. 2(c).

After these phase nonuniformities were removed with the liquid gate, the device was successfully used as the input stage for an optical correlator.^{4,5}

In conclusion, our experiments show that the LCTV can be a highly versatile SLM when used in conjunction with a liquid gate, as described above. Moreover, the device is very inexpensive and can be easily interfaced.

3. OPTICAL IMPLEMENTATION OF A PROGRAMMABLE ASSOCIATIVE RECALL SYSTEM USING LCTV SLMs

The algorithm and architecture of an inner product associative retrieval technique have been published previously.⁶ This system allows an input vector to be compared with M stored test vectors.

An optical system for implementing the inner product associative-retrieval model is shown in Fig. 3. In the figure, the LCTV as described above is used as an electronically addressed SLM. Two multiple-focus holographic lenses (hololenses),⁷ MHL1 and MHL2, are used as space-variant optical elements for performing the necessary inner product and summation functions. The function or capability of the MHLs is shown in Fig. 4. The holographically made optical element is capable of replicating an input 2-D image by first focusing it onto an array of focal points at its focal plane and then forming an array of images at its image plane. All of the M replicated images are derived from a single image through the common aperture of the MHL, a unique capability that cannot be achieved with a refractive element such as a glass lens.

To set up the optical array processing system, we first display the M known vectors on LCTV1, each in a \sqrt{N} by \sqrt{N} matrix format. The M vectors are stored in a \sqrt{M} by \sqrt{M} format. The transmissive screen of the LCTV modulates the input laser light at each spatial location in the two-dimensional screen. Hence, once the writing process is completed, the storage step is done.

To initiate the addressing-recalling mode of operation of the system, an input vector can be invoked. The N-tuple vector should be of the same matrix format as that of any one of the M stored vectors. The screen of LCTV0 can then be used to modulate the collimated laser beam through the cube beamsplitter B0. The information being written on LCTV0 contains one of the M vectors. After the light carrying the vector information is transmitted through B1 and MHL1, it is replicated to a \sqrt{M} by \sqrt{M} (we assume that M is an integer) array of matrices of \sqrt{N} by \sqrt{N} dimensions of elements coinciding matrix by matrix with what has been previously stored in LCTV1. The inner product type of operation is thus partially achieved at this step.

To complete the inner product process, the multiplied results are followed by a diffuser that averages the light intensity and yields the M inner product scalars. The scalars are then transmitted through (i.e., multiplied) to another array of M known vectors displayed (or written on) LCTV2. The matrix arrays on LCTV1 and LCTV2 should be identical. Finally, the weighted vectors (or matrices) are summed by MHL2, which plays a reversed role from MHL1. A TV

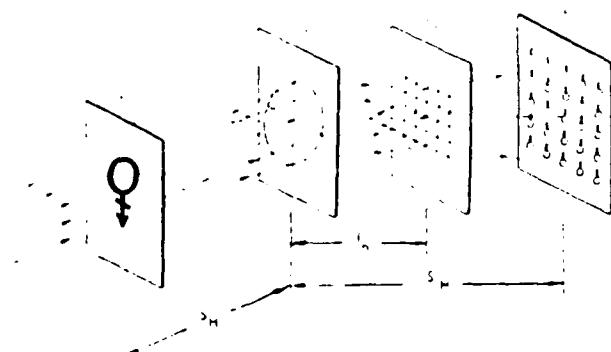


Fig. 4. Function of the multiple-focus holographic lens.

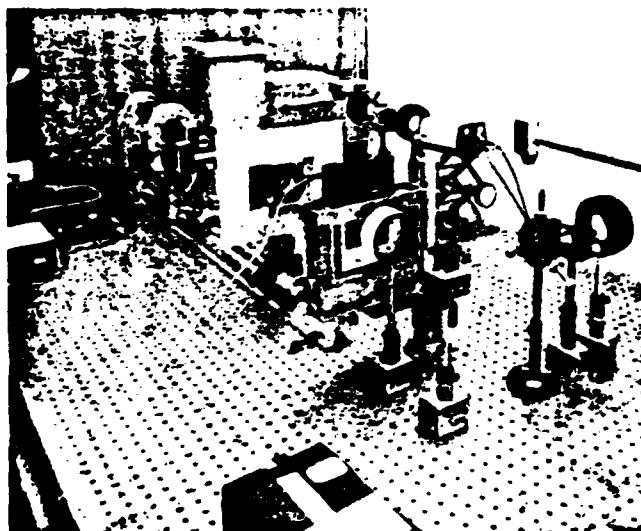


Fig. 5. Photograph of a part of the laboratory setup of the optical inner-product array processor.

camera or CCD detector array can be used to detect the output. The output should be thresholded by an electronic device. (The thresholded array may also be performed by an optical device if a high-gamma spatial light modulator is available.) If the electronic device is adopted (as shown in Fig. 3), the thresholded result can be written on LCTV3 and illuminated by a collimated laser light from B0. This thresholded matrix is then used as a feedback input through B1, and MHL1 is ready for a new iteration process. This procedure will repeat until a steady state is reached and exhibited in the display sent also through B1. Once the steady state is reached, the retrieval process is thus completed.

A partial optical setup of the inner product associative retrieval model is shown in Fig. 5. In the forefront, a LCTV imbedded in a liquid gate is shown. The electronic connection to the LCTV is also visible. The setup is for the purpose of testing the quality of the LCTV in its application as a spatial light modulator. The results of the testing are important to the implementation of the associative retrieval model.

The associative retrieval model becomes a pattern recognition system if one views the M vectors as prestored images of the objects under investigation. Use of the electronically addressable SLM allows computer update of prestored images

4. CONCLUSIONS—RELEVANCE TO THE A&R PROGRAM AT NASA-JPL

The NASA Space Station anticipates automated robots performing progressively unsupervised and unconstrained tasks. This requires 3-D view-independent and context-sensitive scene-understanding capability. The computer systems available today are inadequate because they are limited to simple object recognition in unoccluded and highly structured task environments. Therefore, new expert vision algorithms and architectures are needed that can reason about task-specific knowledge to (1) intelligently control sensor deployment and feature-extraction algorithm parameters, (2) resolve ambiguity in alternative scene interpretations, and (3) create and maintain stable world models useful for path planning and verification.

Optical array processors offer the potential for high computational rates and parallelism in small lightweight packages with low power requirements. The inner-product associative retrieval and compact large memory optical implementation will be useful for the realization of the perceptive vision for telerobots for Space Station applications. As described above, the system has great versatility since both the M stored vectors and the test vector can all be easily and remotely updated. Use of the LCTV allows a convenient way of testing this system.

5. ACKNOWLEDGMENTS

This research is partially supported by NASA and by the Physics Division of the U.S. Army Research Office through contract with the Jet Propulsion Laboratory.

We thank R. A. Lilly for helpful discussions and K. D. Krenz and J. Diep for assistance in the experiments.

6. REFERENCES

1. J. Grinberg, A. Jacobson, W. P. Bieha, L. Miller, L. Fraas, D. Boswell, and G. Myer, "A new real-time noncoherent to coherent light image converter—the hybrid field effect liquid crystal light valve," *Opt. Eng.* 14(3), 217-225 (1975).
2. W. E. Ross, D. Psaltis, and R. H. Anderson, "Two-dimensional magneto-optic spatial light modulator for signal processing," *Opt. Eng.* 22(4), 485-490 (1983).
3. C. Warde, A. M. Weiss, A. D. Fisher, and J. I. Thackara, "Optical information processing characteristics of the microchannel spatial light modulator," *Appl. Opt.* 20, 2066 (1981).
4. H. K. Liu, J. A. Davis, and R. A. Lilly, "Optical data processing properties of a liquid crystal television spatial light modulator," *Opt. Lett.* 10, 635 (1985).
5. J. A. Davis, R. A. Lilly, K. D. Krenz, and H. K. Liu, "Applicability of the liquid crystal television for optical data processing," in *Nonlinear Optics and Applications*, P. Yeh, ed., Proc. SPIE 613 to be published in 1986.
6. S. Y. Kung and H. K. Liu, "An optical inner-product array processor for associative retrieval," in *Nonlinear Optics and Applications*, P. Yeh, ed., Proc. SPIE 613 (to be published in 1986).
7. Y. Z. Liang, D. Zhao, and H. K. Liu, "Multilocus dichromated gelatin hololens," *Appl. Opt.* 22, 3451 (1983).

Hua-Kuang Liu: Biography and photograph appear following "Guest Editorial" (p 807)



Sun-Yuan Kung received his B.S. degree in 1971 from the National Taiwan University, the M.S. in 1974 from the University of Rochester, and the Ph.D. in 1977 from Stanford University, all in electrical engineering.

In 1974 he joined Amdahl Corporation Sunnyvale, Calif., as an associate engineer in LSI design and simulation. From 1974 to 1977 he was a research assistant of Information Systems Laboratories, Stanford University. Since 1977 he has been on the faculty of Electrical Engineering—Systems at the University of Southern California where he is presently an associate professor. In 1984 he was a visiting professor at Stanford University, and later in the same year a visiting professor at Delft University of Technology, the Netherlands.

Dr. Kung's research interests are in the areas of approximation theory in linear systems, digital signal processing, modern and high resolution spectrum analysis, parallel array processors, and VLSI supercomputing for signal processing. Since 1981 he has been in charge of the ONR Selected Research Project (SRO-II) on the development of massively parallel signal processors. He served as the general chairman of the USC Workshop on VLSI Signal Processing (Los Angeles 1982), was a U.S. delegate to the U.S.-Japan Joint Seminar on Mathematical Systems Theory (Gainesville, Fla., 1983), and was an editor of the advanced research book *VLSI and Modern Signal Processing* (Prentice-Hall, 1984). He currently serves on the IEEE ASSP Technical Committee on VLSI and is the associate editor for the VLSI area for *IEEE Transactions on Acoustics, Speech and Signal Processing*. Dr. Kung is a member of ACM and a senior member of IEEE.



Jeffrey A. Davis was born in Meriden, Connecticut, on July 26, 1943. He received the B.S. degree in physics from Rensselaer Polytechnic Institute in 1965 and the Ph.D. in physics from Cornell University in 1970. He is currently a professor of physics at San Diego State University, where he is director of the electro-optics program. His research interests include optical pattern recognition, spatial light modulators, and optical data processing.

INNER-PRODUCT OPTICAL NEURAL PROCESSING AND SUPERVISED LEARNING.

Hua-Kuang Liu, Tien-Hsin Chao and Jacob Barhen California Institute of Technology/Jet Propulsion Laboratory, Pasadena, CA 91109 and Graf Bittner Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093

It has been discovered that an inner-product approach [1.2] to artifical neural network (ANN) computations is not only very suitable for optical implementation but also superior algorithmically. This is because, for a large class of applications, the storage space needed is about an order of magnitude smaller compared to that of the outer-product approach. Moreover, the pre-calculation of the outer-product interconnection matrix is not needed. Hence there is no saturation of the synapses problem in the hardware associated with the interconnection matrix.

For a compairson between the inner- and outer-product approaches with regard to speed of convergence and false states, we have used computer simulation to study the corresponding basins of attraction. As a typical example, in an exhaustive search mode, we have limited our number of neurons to 16 and thus the number of input states is 65536. We have randomly selected up to 4 of the 65536 possible vectors and stored them in the memory and used all 65536 states to associatively retrieve the output both with the inner- and outer-product approaches. We have made the diagonal terms of the outer-product equal to zero and not-equal to zero. The computations have been performed for both using the unipolar and bipolar vectors. In the unipolar case, an adaptative thresholding method is used. The statistical results show the histogram of stable states with respect to the Hamming distances between the input and output and convergence of the stabalized output versus the number of iterations. Based on these preliminary results, it seems that the inner-product approach may have a lesser number of spurious states and converge faster. The other characteristics are similar for the two approaches. We have also investigated the aspect of learning in the inner-product approach. A supervised learning scheme can be realized in the inner-product approach by directly adding or removing vectors from the memory or by revising the memory. This can also be done conveniently in our optical implementations. The synapses that are not explicitly shown in the inner-product approach may be considered as hidden in the architecture. The parallelism and speed-of-light multiplication and summation advantages are fully utilized in this approach of implementing the ANN.

References:

1. S. Y. Kung and H. K. Liu, "An Optical Inner-Product Array Processor for Associative Retrieval," Non-linear Optics and Applications, P. Yeh, Ed., Proc. SPIE, 613 (1986).
2. Hua-Kuang Liu, S. Y. Kung, and Jeffrey A. Davis, "Real-time Optical Associative Retrieval Technique," Opt. Engr., 25, 853 (1986).

Letters to the Editors should be addressed to the Editor, APPLIED OPTICS, Georgia Institute of Technology, Atlanta, GA 30332-0252. If authors will state in their covering communications whether they expect their institutions to pay the publication charge, publication time should be shortened (for those who do).

Infrared predetection dynamic range compression via photorefractive crystals

Hua-Kuang Liu and Li-Jen Cheng

California Institute of Technology, Jet Propulsion Laboratory, Pasadena, California 91109.

Received 22 June 1987.

0003-6935/88/061006-02\$02.00/0.

© 1988 Optical Society of America.

Predetection dynamic range compression is important in solving the problem where the input image has such a high dynamic range that no detectors or sensors can record the complete intensity range of the image without saturation. An example is in the detection of a scene with a shiny automobile or an aluminum building under strong solar illumination in a background of low reflectance. Similar IR scenes of military or industrial interests also pose a problem for image recording and/or data acquisition. To our knowledge, no other means for solving this problem, including the one proposed here,¹ has been reported in the literature.

In this Letter, we describe a new concept using the photorefractive crystals for real-time predetection dynamic range compression. This idea, in principle, is opposite to the idea of using photorefractive crystals for amplification.² In our idea, the nonlinear dependence of the gain coefficient on intensity is not only desirable but also necessary. Experimental results from a study of the two-wave mixing in GaAs crystals³ at 1.15 μm are presented to demonstrate the feasibility of the idea.

First, we present the basic idea of predetection dynamic range compression. The dynamic range (DR) of an input image may be defined as follows:

$$\text{DR} = I_{\max}/I_{\min}. \quad (1)$$

where I_{\max} and I_{\min} represent the maximum and minimum intensities of the input image. Before the image is received at the detector, a device may be used to map the input image into an output image in the following functional form:

$$I_{\text{out}} = f(I_{\text{in}}). \quad (2)$$

where f is a function of mapping, I_{in} and I_{out} are the input and output intensities, respectively. All the intensities may be in 1- or 2-D form. The function f represents the operational characteristics of the device. For example, we may write

$$I_{\text{out}} = A I_{\text{in}}^x,$$

where A is a constant and x is an index number. When $x = 1$ and $A < 1$, the device acts as a linear absorber. When $x = 1$ and $A > 1$, the device acts as a linear amplifier. When $x > 1$, it acts as a nonlinear amplifier. When $x < 1$, it can achieve DR compression, because intensity reduction at brighter spots is larger than that at weaker ones. The essence of DR compression is to find a device with a characteristic f so that f will cause

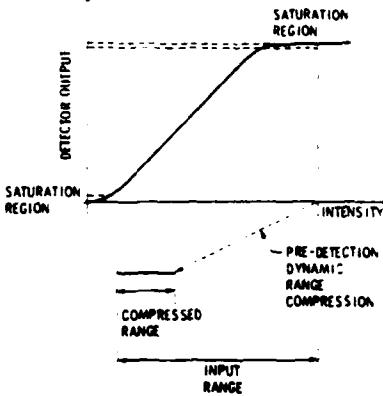


Fig. 1. Predetection dynamic range compression. The upper part of the figure shows the characteristic of a detector or a sensor with the original range of inputs and its saturation effect. The compressed range of the output is shown in the lower part.

$$\text{DR}(\text{output}) < \text{DR}(\text{input}). \quad (3)$$

The idea and its usefulness can be illustrated with the assistance of Fig. 1. The upper part of the figure shows the characteristic of a typical detector, such as a photographic film with a limited detection range. If the input intensity range is too large for the detector's detection range, saturation occurs as shown. After predetection DR compression, this saturation problem is avoided. Since there is still a one-to-one correspondence in terms of the input details, the information capacity is kept intact in this process.

Next, we describe how the two-wave mixing (beam coupling) scheme can be used to compress the DR of an input image. It is well known that during the two-wave mixing process in a photorefractive crystal, one beam can gain energy at the expense of the other.⁴ This effect can be used for DR compression provided the fraction of energy transferred increases with intensity. This particular phenomenon was recently observed in our experiment measuring intensity dependence of the beam coupling coefficient in GaAs.³ The result shows that the measured coefficient increases approximately with the logarithm of beam intensity within the experimental intensity range.

A conceptual DR compression device utilizing the photorefractive effect is shown in Fig. 2. The input beam I_{in} is split into two equal intensity beams, $I_{\text{in}1}$ and $I_{\text{in}2}$, at the beam splitter (BS). Then the two beams exchange energy through two-wave mixing in the GaAs crystal. The output beam that lost energy is represented by $I_{\text{out}1}$, and the one that gained energy is represented by $I_{\text{out}2}$. It can be shown⁴ that the gain coefficient Γ may be written as

$$\Gamma = (1/Z) \ln[(I_{\text{in}1} I_{\text{out}2})/(I_{\text{in}2} I_{\text{out}1})], \quad (4)$$

where Z is the crystal thickness. Assuming $I_{\text{in}1} = I_{\text{in}2} = I_{\text{in}}/2$, we obtain

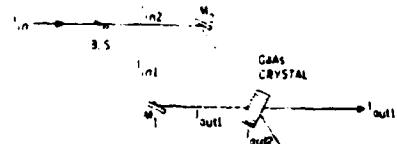


Fig. 2. Diagram depicting a conceptual DR compression device using two-wave mixing in a photorefractive crystal.

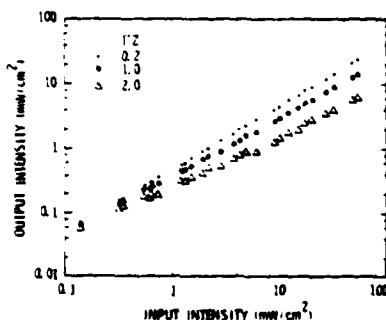


Fig. 3. Experimental data demonstrating the feasibility of DR compression relationship based on Eq. (5) and the GaAs experimental data.

$$I_{out1} = [I_{in1}/[1 + \exp(-\Gamma Z)]]\exp(-\Gamma Z). \quad (5)$$

The results of experiments by Cheng and Partovi³ have revealed that Γ increases with logarithm of beam intensity and the rate of the increase reduces when I_{in1} and I_{in2} are larger than 10 mW/cm^2 . Using the experimental data, we have calculated the output values of a conceptual GaAs photorefractive DR compressor based on Eq. (5). The results, as shown in Fig. 3, clearly indicate that DR compression occurs. The effect is more revealing when the product of the gain and crystal thickness ΓZ is greater. For example, the compression factor is ~ 10 when the product is ~ 2 ; this is achievable by applying an electric field.

In our experiment, the maximum measured value of Γ was only $\sim 0.21 \text{ cm}^{-1}$ at 100 mW/cm^2 of a $1.15\text{-}\mu\text{m}$ laser beam. However, if an electric field was applied to the GaAs, a λ value of 0.9 cm^{-1} was reported.⁵ Further improvement to $\sim 1 \text{ cm}^{-1}$ is still possible.⁶ With the improved gain coefficient, a crystal of 0.3-cm thickness that is viable for a device could offer significant DR compression. However, an effective demonstration of the proposed dynamic range reduction concept using the photorefractive effect has to be delayed until either a successful achievement in the coupling coefficient improvement in GaAs or a discovery of a new photorefractive material having a similar nonlinear property but with a large coupling coefficient.

It should be noted that light absorption in the crystal is neglected in Eq. (5). This does not affect the purpose of this Letter, the demonstration of the DR compression concept. However, light absorption can reduce I_{out1} by a constant factor.

The research described in this paper was performed at the Jet Propulsion Laboratory and jointly supported by the Defense Advanced Research Projects Agency, the Strategic Defense Initiative Organization, the Physics Division of the U.S. Army Research Office, and the National Aeronautics and Space Administration.

References

- H.-K. Liu and L.-J. Cheng, "Real-Time Dynamic Range Compression," *J. Opt. Soc. Am. A* 3(13), P 51 (1986).
- Y. Fainman, C. C. Guest, and S. H. Lee, "Optical Digital Logic Operations by Two Beam Coupling in Photorefractive Material," *Appl. Opt.* 25, 1599 (1986).
- L. J. Cheng and A. Partovi, "Temperature and Intensity Dependence of Photorefractive Effect in GaAs," *Appl. Phys. Lett.* 49, 1456 (1986).
- N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic Storage in Electrooptic Crystals. I. Steady State and II. Beam Coupling—Light Amplification," *Ferroelectrics* 22, 949 (1979); 22, 961 (1979).
- G. Albanese, J. Kumar, and W. H. Steier, "Investigation of the Photorefractive Behavior of Chrome Doped GaAs Using Two-Wave Coupling," *Opt. Lett.* 11, 650 (1986).
- G. C. Valley, "Two-Wave Mixing with an Applied Field and a Moving Grating," *J. Opt. Soc. Am. B* 1, 868 (1984).

Stereoscopic uses of diffraction gratings

Marcelo Trivi and Hector J. Rabal

Centro de Investigaciones Ópticas, CIOP, Casilla de Correo 124, 1900 La Plata, Argentina.

Received 10 June 1987.

0003-6935/88/061007-03\$02.00/0.

© 1988 Optical Society of America.

This Letter concerns the imaging of 3-D objects in optical systems using diffraction gratings.

It can easily be seen experimentally that when one observes a 3-D scene through a diffraction grating, the objects appear symmetrically replicated with respect to the ordinary (zero-order) image. As chromatic dispersion affects the image, a monochromatic source is used. Careful observation shows that closer objects produce twin images that are less separated than far away objects, and separation as a function of the distance reaches saturation (i.e., the twin images produced by the stars do not seem to be separated much more than those due to an object several meters away from the observer).

This experiment suggests that by calibrating the distances between diffracted orders with the actual distances to the corresponding objects, a simple telemeter could be designed. It would be so, at least in principle, but as a practical instrument it would have several shortcomings. For example, for more or less accurate measurement of the distances between points of the twin images, the object should have abrupt luminance discontinuities. Besides, such distances change with eye-grating distance, complicating the measurement even more.

Figure 1 is a photograph of a straight line luminous object, the points of which recede from the observer, as seen through a diffraction grating. Only the zero and the first diffracted orders are shown. It can be seen that the distance between homologous points on the diffracted orders to the zero order increases with the distance to the observer in approximately linear fashion.

A careful analysis involving diffraction theory shows that the image of a point source through a system, such as that depicted in Fig. 2, is described by:¹

$$u(x,y) = C \left(\frac{z_{13}}{z_{12}} \right)^2 q \left[x, y; \frac{1}{\lambda z_{14}} \left(1 - \frac{z_{13} z_{23}}{z_{12} z_{14}} \right) \right] \\ \times T \left(\frac{x}{\lambda z_{12} z_{14}}, \frac{z_{13} y}{\lambda z_{12} z_{14}} \right). \quad (1)$$

A Reprint from the

PROCEEDINGS

Of SME-The International Society for Optical Engineering



Volume 638

Hybrid Image Processing

1-2 April 1986
Orlando, Florida

Optical image subtraction techniques, 1975-1985

Hue-Kuang Liu, Tien-Hsin Chao
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California 91109

©1986 by the Society of Photo-Optical Instrumentation Engineers
Box 10, Bellingham, Washington 98227 USA Telephone 206/678-3280

Optical Image Subtraction Techniques, 1975 - 1985

Hua-Kuang Liu and Tien-Hsin Chao

Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California 91109

Abstract

A brief review of optical image subtraction techniques recently developed is presented. We have divided them into two categories according to whether they can or cannot be performed in real time.

Introduction

Optical image subtraction (OIS) techniques¹ are very useful in many areas such as earth resources studies, meteorology, automatic surveillance, pattern recognition, urban growth studies, and bandwidth compression in communication systems. An earlier review paper written by John F. Ebersole² has covered many optical subtraction techniques up to 1975. It is the purpose of this paper to present a brief review of many OIS techniques developed since then.

From an application standpoint, we classify the OIS techniques into two groups, depending on whether or not they can be operated in real time. A Table listing the techniques, with proper references, is given for the reader's convenience. Because of the extensive amount of literature on OIS, we might not have included them all in this article.

Real-time OIS

The real-time OIS techniques are generally realized through the holographic and interferometric methods. In an experimental process typically involving the use of two stored scenes recorded on either photographic films or real-time recording devices such as a $\text{Bi}_{12}\text{GeO}_{20}$ (BGO) crystal, the input objects are illuminated by either a coherent or incoherent light source. Shearing devices such as gratings, Wollaston prisms, and holographic shear lenses may be used to shear the two input objects and recombine them in the output plane. A π -radian phase change is introduced between the two light paths passed through by the two output images and thus the optical interference can eliminate the common background, and only the difference image will be seen in the output plane.

Source encoding

A source encoding technique will perform image subtraction with an extended quasi-monochromatic light source,³ as shown in Figure 1. A 1-D mask consisting of a periodic array of narrow slits with slit width d and nearest neighbor separation D is used to create a high degree of spatial coherence between any pair of points (at the input plane P_2) that are separated by $\Delta x = n(\lambda f/D)$, where n is an integer, λ is the wavelength of the light source, and f is the focal length of the transform lenses. Following the Van-Cittert Zernike theorem, the spatial coherence function in the input plane is the normalized Fourier transform of the mask intensity transmittance function as shown in Figure 2. If h_0 is selected to be $\lambda f/D$, and the ratio of d/D is $1/10$, the spatial coherence of any point-pair separated by a distance of $2h_0$ is about 0.936. The spatial coherence can further be increased through the use of a narrower slit. A sinusoidal diffraction grating is inserted into the Fourier plane P_3 . The grating period is chosen to match that of the encoding mask. The grating will recombine the input image transparencies S_1 and S_2 at the output plane P_4 with a π -phase difference. Thus a high quality difference image can be obtained.

This source encoding technique can also be extended to white-light image subtraction.⁶ As shown in Figure 3, a fan-shaped encoding mask is illuminated by a rainbow spectrum dispersed by a volume grating using a white-light extended source. The scale of the fan-shaped mask varies linearly to accommodate the wavelength variation. The diffraction grating used in the Fourier plane is also modified into a matching fan-shape. This white-light image subtraction system is particularly suitable for color signal processing.

Polarization modulation⁸

The basic principle of polarization modulation OIS may be illustrated in Figure 4(a). A coherent plane wave is utilized to illuminate the two input objects S_1 and S_2 through the polarizers P_1 and P_2 , respectively. S_1 and S_2 are separated by a distance x_0 from the optical axis. As a result of polarization modulation, the polarization of the distribution

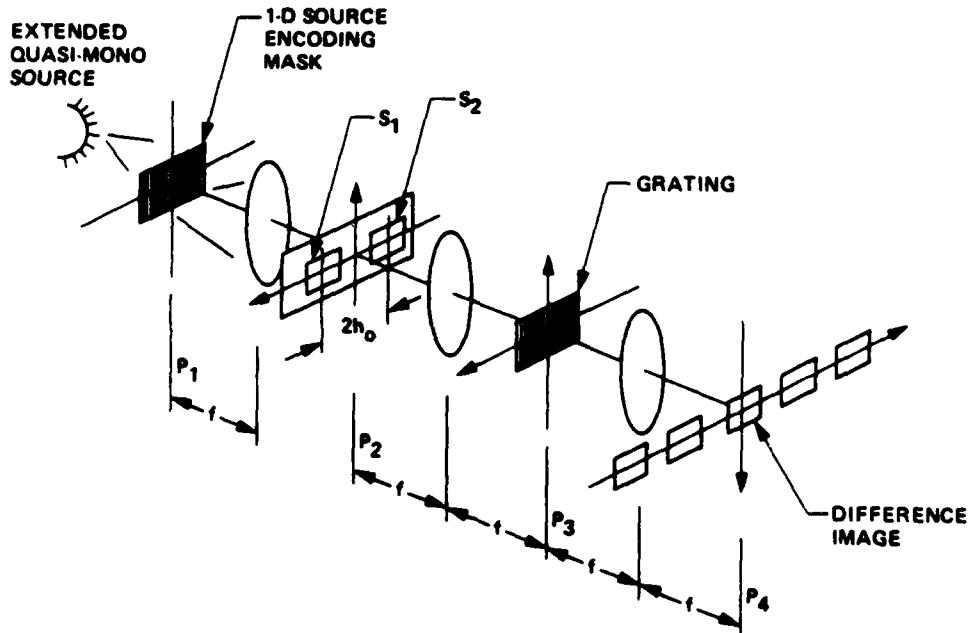


Figure 1. A diagram showing image subtraction by encoding an extended source

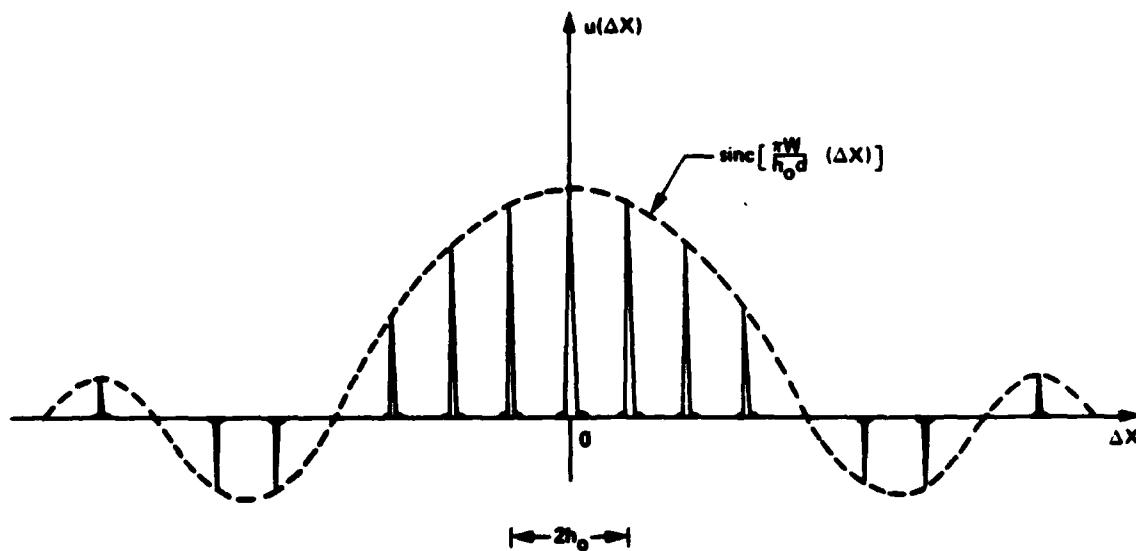


Figure 2. Spatial coherence function created by a source encoding mask

immediately behind S_1 and S_2 state is at a 90° difference. At the back focal plane of the Fourier transform lens L_1 , a birefringent Wollaston prism is inserted which deflects the light beams passing through it according to the polarization state. The deflection angles are equal but at opposite direction such that the image of S_1 and S_2 can be recombined at the output plane. A rotatable analyzer is placed in front of lens L_2 . As the analyzer is adjusted by a crossed-Nicol geometry, as shown in the vector diagram of Figure 4(b), π -phase shift between S_1 and S_2 can be obtained and hence image subtraction can be achieved.

Pseudocolor image-difference detection technique

The polarization modulation technique can be extended for detecting image differences through pseudocolor encoding^{9,10}, as shown in Figure 5. A pair of input objects S_1 and S_2

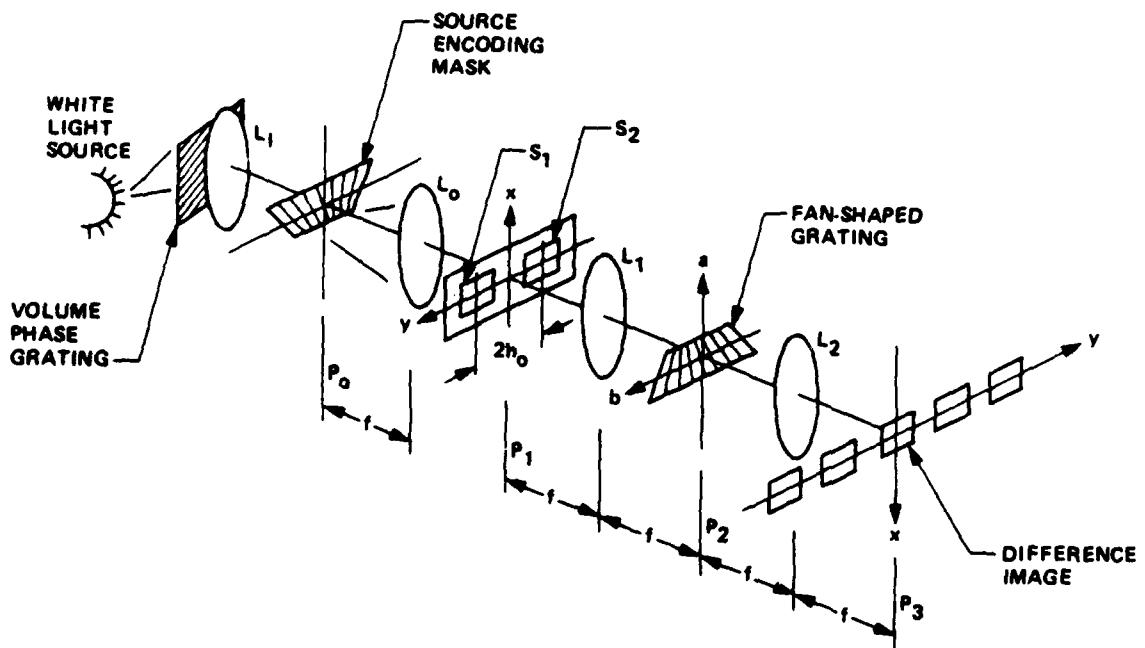


Figure 3. A white-light image subtraction processor using a source encoding technique

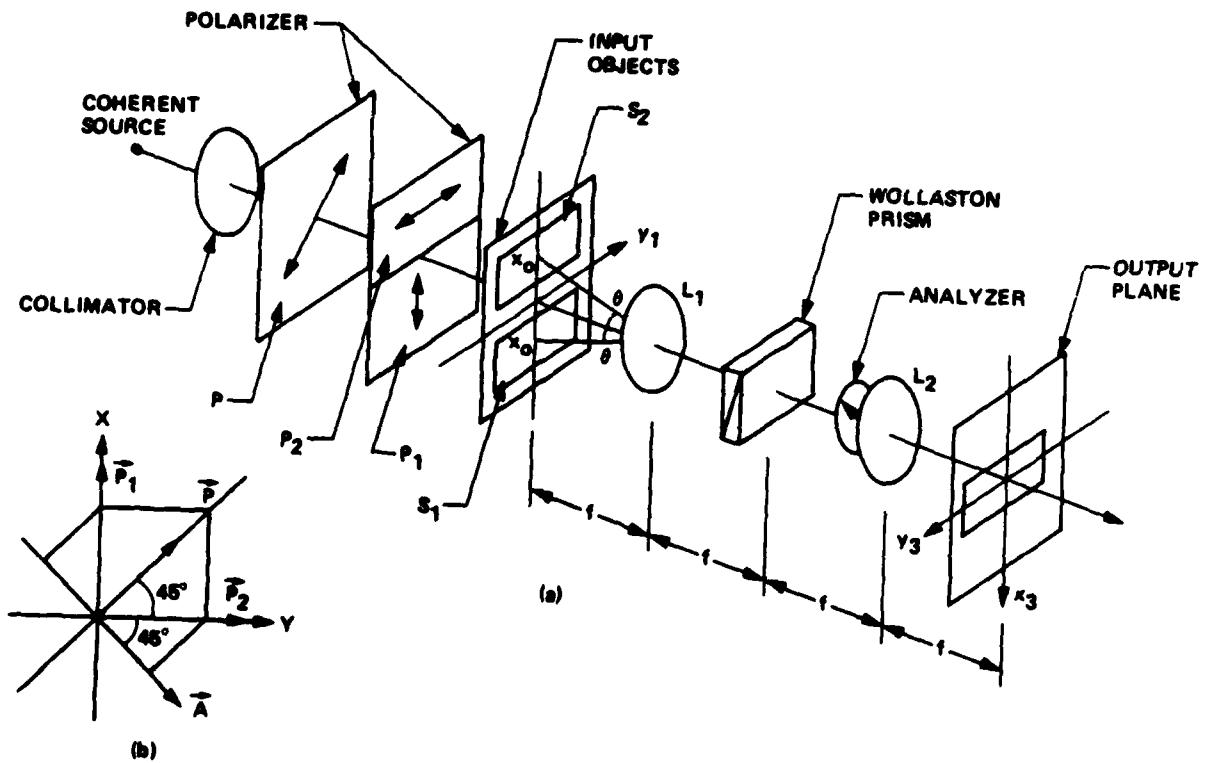


Figure 4. A polarization modulation OIS system

are inserted into the input plane P_1 and are illuminated by collimated white-light. In this case, S_1 is covered by a green filter and S_2 by a red filter. A Wollaston prism is inserted behind lens L . With proper alignment, the prism will combine the two separate input images at the output plane P_2 . The combination forms a synthetic pseudocolor image with its color

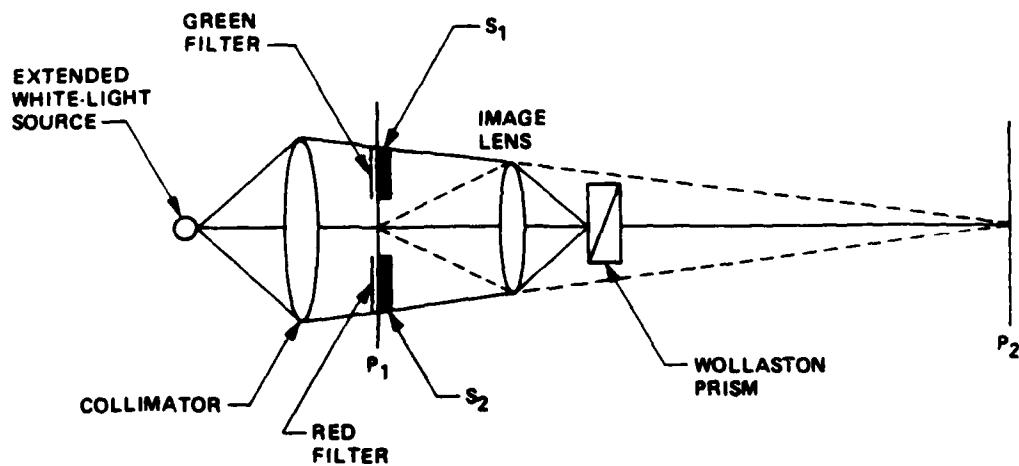


Figure 5. Real-time white-light image-difference detection system

indicating the difference between S_1 and S_2 . In the present case, the red color shows the portion where the transmittance S_2 is greater than that of S_1 and vice versa for the green. The yellow shows the identical portion of S_1 and S_2 . If a colorimetric detector is used to scan the output, the differences in the image can be found.¹⁰

Holographic shear lens technique

The holographic shear lens OIS technique¹¹ is a modification of the diffraction grating method³. As shown in Figure 6, two photographic image transparencies S_1 and S_2 are inserted into the input plane of the coherent image subtraction system. A specially fabricated holographic shear lens is inserted into the Fourier plane. The holographic lens shears and recombines the images S_1 and S_2 at the output plane. The holographic shear lens is synthesized by the superposition of two Fresnel zone plates with carrier frequencies separated by Δv . This shear lens is similar to the double frequency grating used in shearing-interferometry. In performing the image subtraction, the shear lens is tilted at a suitable angle such that the light path traveled by the two input images is differed by $\pi/2$ and so that the subtraction between the image can be obtained. A similar approach¹² utilizing a double-grating shearing interferometer to shear, and introducing π -radian phase change between a pair of input transparencies has also been proposed.

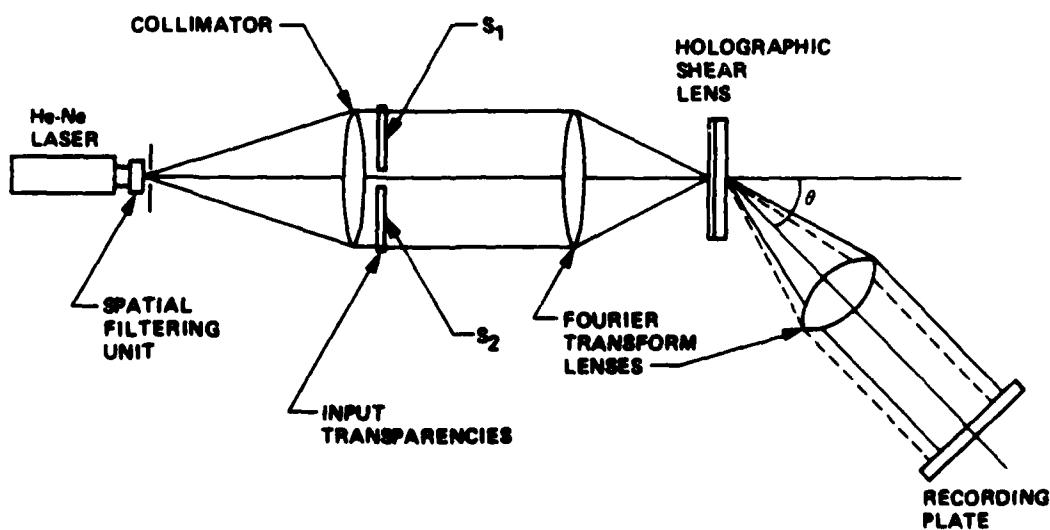


Figure 6. A set-up of real-time image subtraction system using holographic shear lens

Nonlinear optics

Real-time OIS can also be achieved by holographic two-wave or four-wave mixing (FWM) techniques using nonlinear electro-optics crystals such as the photorefractive $\text{Bi}_{12}\text{GeO}_20$ (BGO) crystal¹³. The set-up of a FWM through a reflection hologram for image subtraction is shown in Figure 7(a). The reflection hologram is formed by recording the interference pattern between the object beam A_o and reference beam A_r that are on the opposite side of the BGO crystal. The crystallographic axis normal to the crystal surface is oriented parallel to the bisector of the reference beam and the object beam. A periodic space-charge electric field is generated which modulates the refractive index of the crystal. A phase volume reflection hologram is formed inside the crystal. The reading beam A_1 can then reconstruct the object beam A_o , which travels to the opposite direction of A_o . A beam splitter is inserted into the light path of A_p and forms a real image of the object that can be picked up by a TV camera and displayed on a TV monitor.

Application of this technique is similar to that of a double-exposure holographic subtraction technique; a phase-shift is introduced into the reference beam A_r by translating the piezomirror a distance $\lambda/4$ between the recording of the first and second holographic images. Due to the slow response and the erase time of BGO as shown in Figure 7(b), the two images can both exist for a short time interval τ , during which the difference image can be shown on the monitor screen.

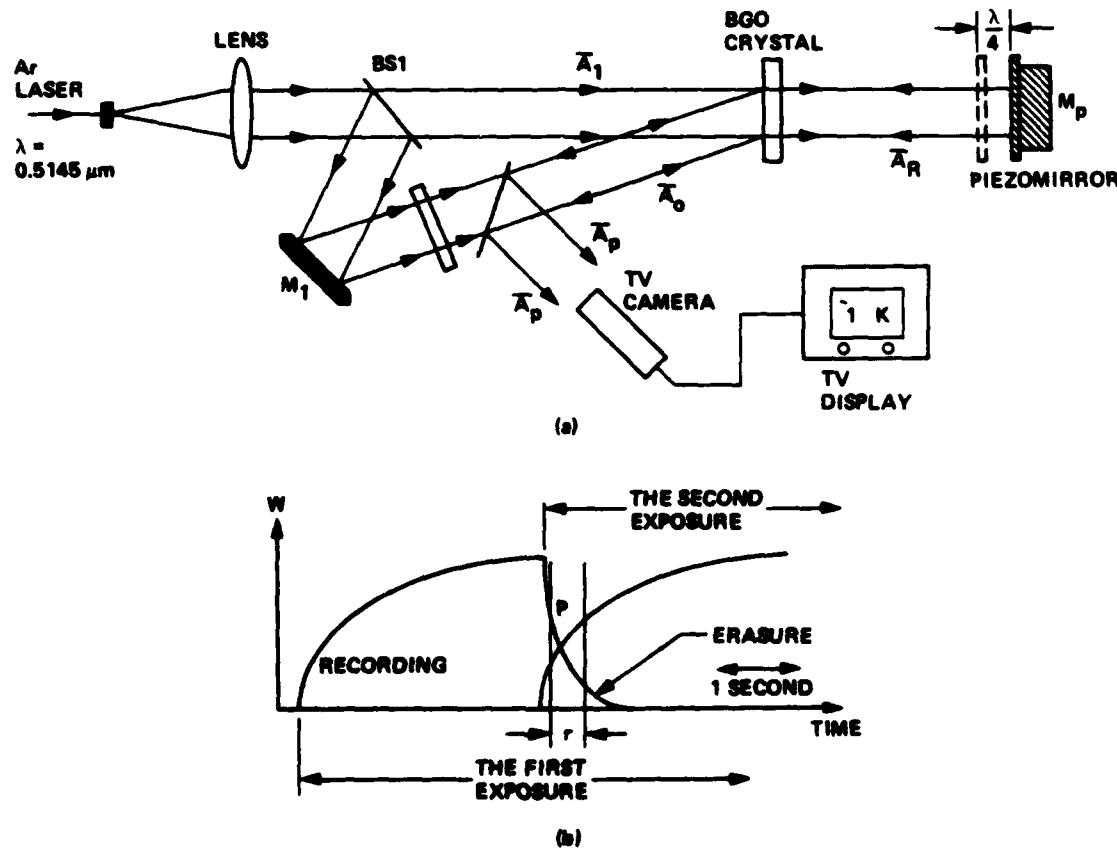


Figure 7. Real-time image subtraction in four-wave mixing with photorefractive $\text{Bi}_{12}\text{GeO}_20$ crystal
 (a) System schematic diagram
 (b) Recording-erasure cycle of holographic recording in BGO crystal

Other electro-optics devices that have been used in optical image subtraction include the Hughes Liquid Crystal Light Valve (LCLV)¹⁴, Pockel Cell¹⁵ and channel plate image intensifier.¹⁶

Non-real-time OIS

Non-real-time OIS techniques are generally achieved through optical encoding. The image transparency S_1 is imaged through an encoding mechanism which modulates the output image of S_1 by a 1-D (possibly periodic) pattern. The coding mechanism is then aligned such that the modulating pattern is shifted by half a period and an exposure of the second transparency S_2 is taken. For example, if the coding pattern is a cosinusoidal grating, the resultant coded image transmittance $t(x,y)$ of the developed film may be written as follows:

$$T(x,y) = S_1(x,y)[1+\cos(2\pi f_0 x)] \\ + S_2(x,y)[1-\cos(2\pi f_0 x)],$$

where f_0 is the spatial frequency of the grating. The coded film is then placed at the image plane of a 4-f optical processor and a band-pass binary filter is placed at the Fourier plane. The first order of the image spectrum can be selected and the difference image therefore appears at the output plane.

Speckle diffuser encoding

This method utilizes the multiple imaging effect of a photographic diffuser¹⁷ produced by photographic recording of a fine speckle pattern with the optical set-up shown in Figure 8. The very coarse grid (Ronchi ruling) G is illuminated by a quasi-monochromatic diffused light using a tungsten lamp S in conjunction with a bandpass spectral filter F and a ground glass. The random diffuser is placed at a distance z_1 from the grating, and the imaging plane P_2 is located at a distance z_2 from the diffuser. The diffuser screen acts like a two-dimensional random array of pinholes with each pinhole producing its own discrete image. The magnification factor $M=z_2/z_1$ has typical values on the order of 2×10^{-4} . During encoding, an input photographic transparency is placed at the multiple imaging plane P_2 in the output plane. The diffuser screen modulates the input transparency with multiple demagnified images. A relay lens is utilized to project the encoded image in plane P_2 to the output photographic plate. Between the two exposures, the coarse grid is translated in its plane by half a period. The extremely high demagnification effect of the speckle diffuser allows the use of a very coarse grating, and the accuracy of the π -phase shift can be increased by a factor equal to the demagnification of the imaging process.

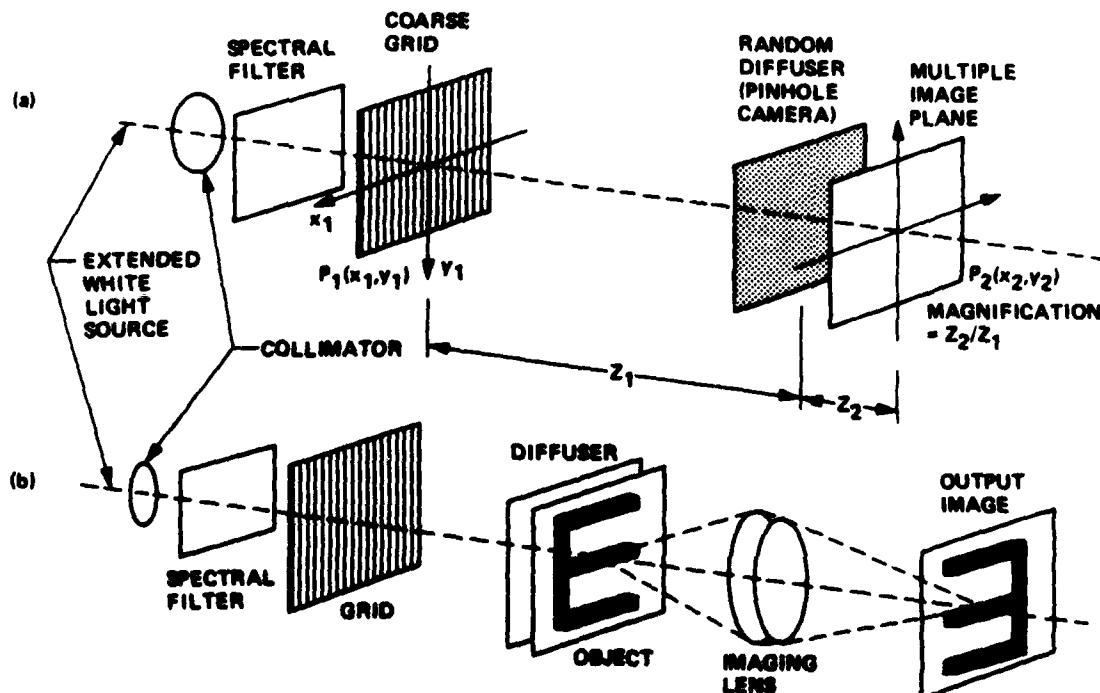


Figure 8. Image subtraction system through speckle diffuser coding
 (a) Set-up showing multiple imaging by diffuser
 (b) Optical recording arrangements

Speckle-pattern encoding

The speckle pattern coding technique is also a double-exposure approach^{18,19} which is suitable for extracting the differences between both 2- and 3-dimensional scenes. Figure 9 shows the set-up used for obtaining the differences between two scenes. A 3-D diffusive object (or a 2-D transparency placed behind a fine diffuser) is illuminated by a monochromatic light. The imaging lens L projects the 3-D scene onto the output photographic plane through an encoding mask which consists of two circular apertures. The output images are encoded with speckle-modulated Young's fringes because of the limited size of the pair of circular apertures. The π -phase shift of the output Young's fringes between two exposures was obtained by inserting a parallel glass plate between the encoding mask and the

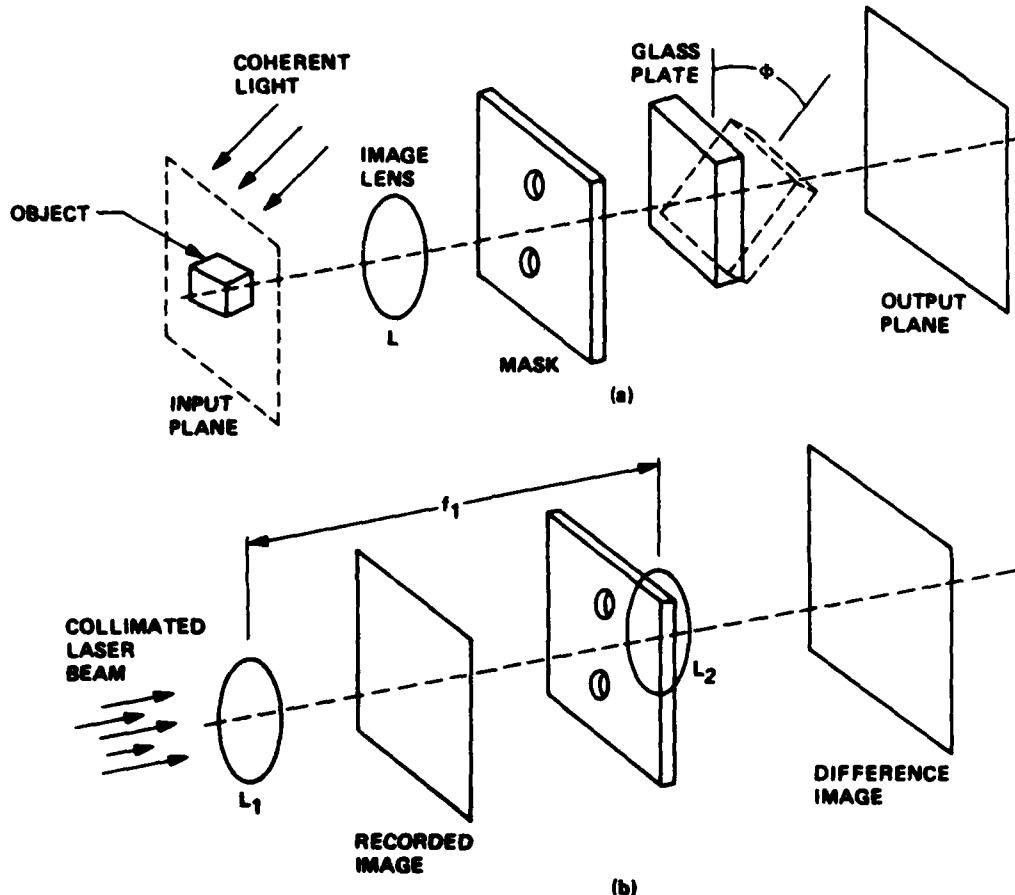


Figure 9. Optical image subtraction through speckle pattern modulated by Young's fringes
 (a) Double-exposure recording set-up
 (b) Reconstruction set-up

output plane. During the second exposure, the plate is adjusted by a pre-determined angle ϕ from its position at the first exposure. The image difference can be obtained by spatial filtering of the encoded photographic image.

Halftone screen encoding

The halftone screen encoding method^{20,21} is a three-step method. The first step is to prepare the halftone negatives S_{1n} and S_{2n} for the input photographic transparencies S_1 and S_2 , respectively. Each of the two images is contact-printed through a 1-D halftone contact screen of spatial period X onto a high-contrast negative film. During the preparation of the halftone negatives S_{1n} and S_{2n} the exposure conditions are the same, and the contact screen is fixed in a position relative to both scenes S_1 and S_2 . The second step is to prepare the halftone positives S_{1p} and S_{2p} which can be done by direct-contact-printing S_{1n} and S_{2n} onto fresh high-gamma negatives.

The intensity transmittance profile in a unit cell X of S_{1n} , S_{2n} , S_{1p} , and S_{2p} represented by T_{S1n} , T_{S2n} , T_{S1p} , and T_{S2p} are shown in Figure 10. The value of T_m is

usually greater than 0.85. W_{S1N} and W_{S2N} are the widths of the opaque bars, and W_{S1P} and W_{S2P} are the widths of the transparent bars W_{S1N} , W_{S1P} , and W_{S2N} , W_{S2P} . Generally speaking, because of the thresholding effect of the high-gamma film, the relationship between W_{S1} and T_{S1} and W_{S2} and T_{S2} can be described as

$$W_{S1}/X = f(T_{S1}), \quad W_{S2}/X = f(T_{S2}).$$

Finally, the subtraction $f(T_{S1}) - f(T_{S2})$ can be obtained by superposing S_{1P} and S_{2N} . Similarly, superposing S_{1N} and S_{2P} can provide the difference $-f(T_{S1}) + f(T_{S2})$. In the case of $f(T_{S1}) - f(T_{S2})$, the center of the opaque bar S_{1N} is made to coincide with that of the transparent spacing of S_{1P} , as shown in Figure 11(a), where either $W_{S1} > W_{S2}$ or $f(T_{S1}) > f(T_{S2})$ is assumed. The transmittance T_s , resulting from the subtractions in the region of the

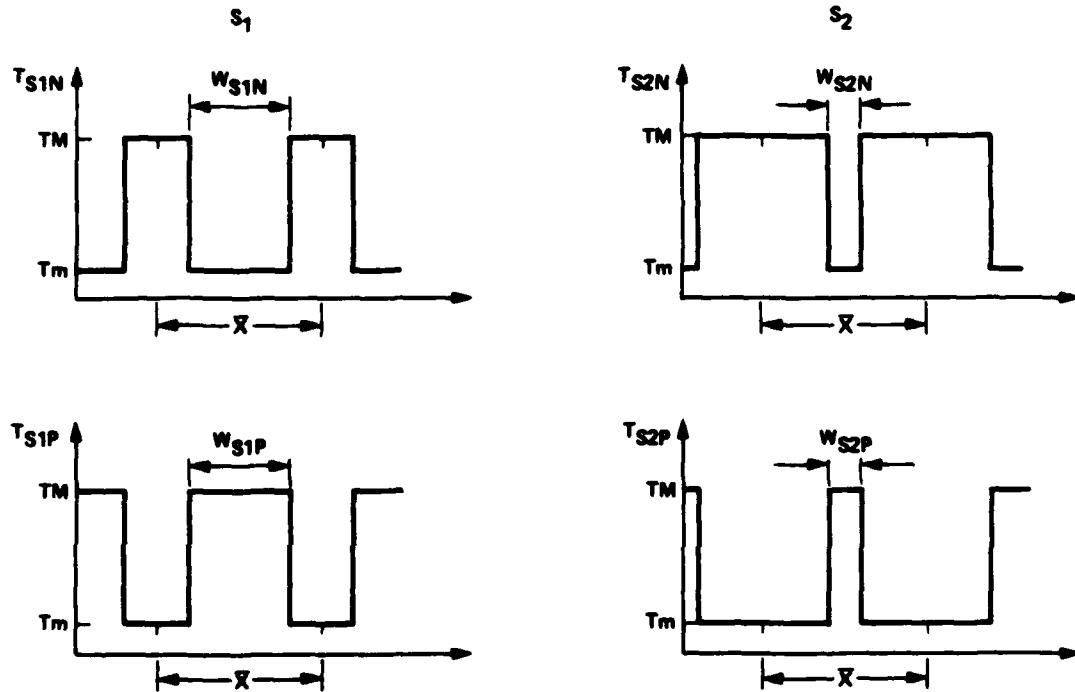


Figure 10. The transmittance profiles of the halftone negatives and halftone positives of two images S_1 and S_2

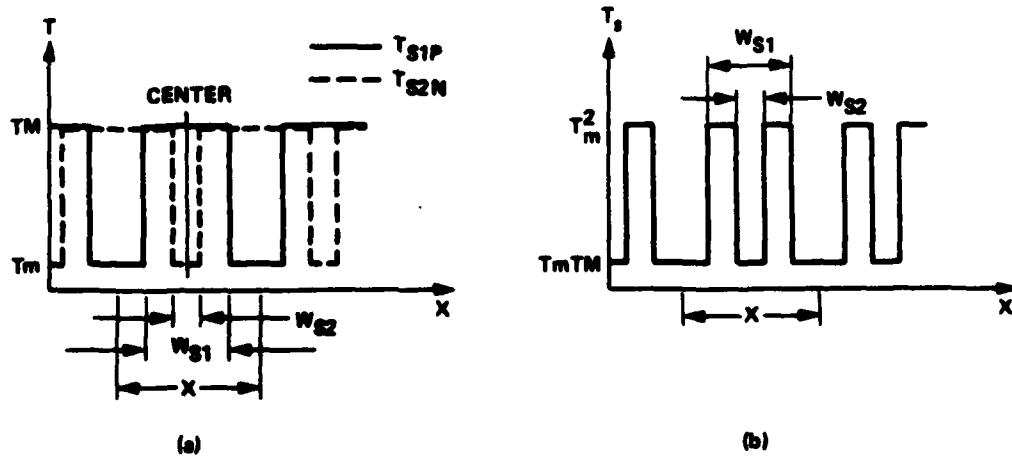


Figure 11. (a) Transmittances of halftone photographs S_{1P} and S_{2N}
(b) Transmittance of S_{1P} superposed on top of S_{2N}

specific cell of the halftone photograph, is shown in Figure 11(b). The subtracted value can be obtained by measuring T_s .

In general, image subtraction can also be achieved through contact-printing encoder ruling using ruling with unequal widths in its opaque and transparent bars. It has been theoretically proven that the use of such a grating does not alter the difference between images.²²

Polarization-shifted encoding

The encoding system is shown in Figure 12(a). A quasi-monochromatic source S is imaged onto a pinhole P by a condensing lens L_1 . The light passing through P is collimated into a plane wave by the lens L_2 . A polarizer P_1 is placed between the sources S and the lens L_1 , and the polarized parallel beam illuminates the grating G. A set of binary polarization filters M is inserted in the back focal plane of the lens L_3 , as shown in Figure 12(b). The filter consists of two mutually orthogonal polarizers, each of which transmits one of the first order of the grating spectrum. The transmission axis of the polarizers in filter M makes angles of $\pm \pi/4$ with that of the polarizer of P_1 . A polarizer P_2 is placed at the

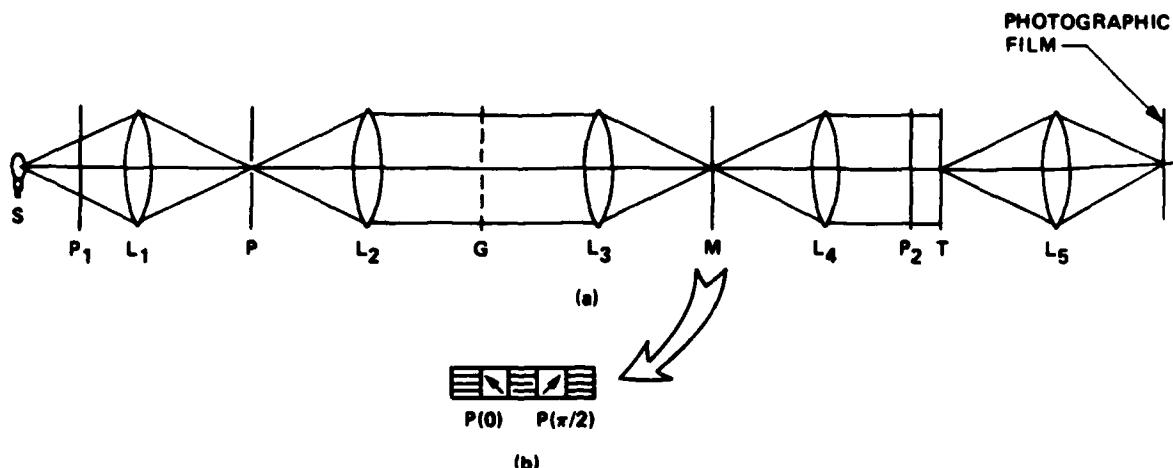


Figure 12. Set-up for recording the modulated images using polarization-shifted spatial carrier
 (a) System set-up
 (b) Binary polarization filter M

output plane of the grating G with its transmission axis inclined to both polarizers of the filter M. A photographic transparency placed immediately behind the polarizers P_2 will be modulated by high-contrast sinusoidal fringes. The modulated transparency is imaged by lens L_5 onto a fresh recording film.

During the second exposure of the object transparencies S_1 and S_2 , the polarizer P_1 is rotated through an angle of $\pi/2$ such that the encoding sinusoidal fringe pattern will shift by a half period. The difference image can then be obtained by spatial filtering.

Conclusion

We have briefly reviewed the state-of-the-art of most of the OIS techniques developed during the last decade. For clarity, the techniques included are listed in Table I. The non-real-time optical techniques are important because some of them may be applied directly and others may be converted into real-time ones through the use of advanced real-time spatial light modulators or electro-optic devices. The most useful techniques should nevertheless be the real-time operations. Since optical techniques can offer true amplitude subtraction in addition to the parallel no-need-of-conversion, high-throughput characteristics, their importance is obvious. Continued development of coherent optical subtraction techniques should hence be of interest to NASA, DoD, and industry.

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We wish to thank C. P. Grover for helpful discussions, along with J. Donato and W. Whitney for editorial help.

Table 1. Optical Image Subtraction Techniques
(1975 - 1985)

Method	Amplitude (A) or Intensity (I) Subtraction	Light Source M: monochromatic Q-M: quasi- monochromatic WH: white light	Point Source: P Slit Source: S Extended Source: E
<u>Real-Time OIS</u>			
a. Source encoding			
1-D encoding mask (Ref. 4,5)	A	Q-M	E
2-D encoding mask (Ref. 6,7)	A	WH	E
b. Polarization modulation (Ref. 8)	A	M	P
c. Pseudocolor image difference detection (Ref. 9,10)	I	WH	E
d. Shearing interferometric methods (Ref. 11,12)	A	M	P
e. Electro-optics method (Ref. 13)	A	M	P
<u>Non-Real-Time OIS</u>			
a. Speckle diffuser encoding (Ref. 17)	I	Q-M	E
b. Speckle pattern encoding (Ref. 18,19)	I	M	P
c. Halftone screen encoding (Ref. 20,22)	I	WH	E
d. Polarization shifted carrier encoding (Ref. 23,24)	I	Q-M	S

References

1. D. Gabor et al., "Optical Image Synthesis by Holographic Fourier Transformation," *Phy. Lett.* **18**, 116 (1965).
2. J. P. Ebersole, "Optical Image Subtraction," *Appl. Opt.*, **14**, 436 (1975).
3. S. H. Lee, S. K. Yao, and A. G. Milines, "Optical Image Synthesis in Real-Time by a Diffraction-Grating Interferometric Method," *J. Opt. Soc. Am.* **60**, 1037 (1970).
4. S. T. Wu and F. T. S. Yu, "Image Subtraction with Encoded Extended Incoherent Source," *Appl. Opt.*, **20**, 4082 (1981).
5. A. A. Lapides, "Image Subtraction by Spatial Filtration Technique," *Optica Acta*, **32**, 329 (1985).
6. F. T. S. Yu, "White Light Optical Signal Processing", Wiley Interscience, New York, 155 (1986).
7. F. T. S. Yu and S. T. Wu, "Color Image Subtraction with Extended Incoherent Sources," *J. Opt.*, **11**, 20 (1982).
8. D. Zhao and H. K. Liu, "Real-Time Optical Interferometric Image Subtraction by Wave Polarization," *Appl. Opt.*, **21**, 3864 (1982).
9. D. Zhao and H. K. Liu, "Real-Time White-Light Pseudocolor Image-Difference Detection Technique," *Opt. Lett.*, **8**, 99 (1983).
10. G. G. Mu, C. K. Chiang and H. K. Liu, "White-Light Image Addition and Subtraction by Colorimetric Measurement," *Opt. Lett.*, **6**, 389 (1981).

11. V. V. Rao, C. Joenathan and R. S. Sirohi, "Real-Time Optical Image Subtraction by a Holographic Shear Lens," *Opt. Comm.*, **55**, 83 (1985).
12. K. Patorski, "Subtraction and Addition of Optical Signals Using a Double-Grating Shearing Interferometer," *Opt. Comm.*, **29**, 13 (1979).
13. Y. H. Ja, "Real-Time Image Subtraction in Four-Wave Mixing with Photorefractive $\text{Bi}_{12}\text{GeO}_{20}$ Crystals," *Opt. Comm.*, **42**, 377 (1982).
14. E. Marom, "Real-Time One-Step Incoherent Level Slicing of Optical Images," *Opt. Comm.*, **26**, 169 (1978).
15. K. M. Bagenbuch and S. P. Sarraf, "Holographic Image Subtraction of High-Velocity Dust Using a Pockels Cell," *Opt. Lett.*, **1**, 325 (1982).
16. A. Seko, "A High-Speed Image Information Preprocessor Using a Channel Plate Image Intensifier," *Opt. Engr.*, **19**, 229 (1980).
17. C. P. Grover, "Optical Image Subtraction Using Multiple-Imaging Effect on Photographic Diffusers," *J. Opt. Soc. Am. A.*, **1**, 145 (1984).
18. H. J. Rabal, M. Bolognini, E. Sicre, and M. Garavaglia, "Optical Image Subtraction Through Speckle Pattern Modulated by Young Fringes," *Opt. Comm.*, **34**, 7 (1980).
19. E. Marom and Z. Kasher, "Optimal Distribution of Multiple Exposure in Speckled Image Subtraction Setups," *J. Optics A*, **5** (1977).
20. D. Zhao, C. C. Chiang and H. K. Liu, "Quantitative Optical Image Subtraction by Means of Pulse-width Modulation Through One-Dimensional Contact Screens," *Opt. Lett.*, **6**, 490 (1980).
21. D. Zhao, C. C. Chiang and H. K. Liu, "Contact Screen Image Subtraction Technique," *Appl. Opt.*, **20**, 4234 (1981).
22. G. Rodriguez, "Grating Coding for Image Subtraction with Uneven Bar Widths," *Opt. Lett.*, **8**, 455 (1983).
23. A. Ghosh and A. K. Chakraborty, "Extraction of Difference Using a Polarization-Shifted Periodic Carrier," *Optica Acta*, **30**, 425 (1983).

APPENDIX E

**New Architectures for
Optical Processing**

Annual Report

A Study on Optical Array Architectures for Image Processing

Submitted to

Dr. H.K. Liu, JPL
by

S.Y. Kung, USC

A Study on Optical Array Architectures for Image Processing

S.Y. Kung

Signal and Image Processing Institute
Dept. Of Electrical Engineering
University of Southern California
Los Angeles CA 90089

August 5, 1987

1 Introduction

With the advent of optical/VLSI technology, memory and processing power is becoming very inexpensive and affordable. The emphasis of optical/VLSI system design is now placed on reducing the overall interconnection complexity and keeping the overall architecture highly regular, parallel, and pipelined. The feasibility of array processors based on optical/VLSI technology enables the processing speed to increase by several orders of magnitude. With optical/VLSI it becomes feasible to construct an array processor which closely resembles the flow graph corresponding to the algorithm of interest. These arrays maximize the strength of optical/VLSI in terms of intensive and pipelined computing and yet circumvent its main limitation on communication. There are four main points of attack in an integrated optical/VLSI system design methodology: *technology, applications, algorithms, and architectures* [18].

In this project, we have adopted a systematic mapping methodology, based on the use of (*data*) dependence graph (DG), to develop a number of array architectures suitable for optical/VLSI (or a hybrid design) for image processing and analysis [18]. The following summarizes several example architectures developed in this project.

Image processing deals with deterministic and stochastic representation of images (image transforms and models), compression of the large amount of data in the images (image data compression), and improving the quality of the image by filtering and by removing any degradations present (image enhancement and restoration). The main areas in digital image processing are *enhancement, restoration, reconstruction and coding*. The derived array architectures in image processing are presented in the following sections:

Section 2: Array architectures for edge detection feature extraction , and for median/rank-order filtering used in image enhancement [12].

Section 3: Array architectures using artificial neural networks (ANNs) for image restoration and pattern recognition [18].

Section 4: Array architectures for Kalman filtering for least squares image restoration/reconstruction [13].

Below a brief introduction to the mapping methodology used to derive the array architectures is given.

1.1 Mapping Algorithms to Broadcasting SFGs

The main concern in algorithm-oriented array processor design is: given an algorithm, *how is an array processor systematically derived?* A fundamental issue is *how to express parallel algorithms in a notation which is easy to understand by humans and possible to compile into efficient optical/VLSI array processors*. The ultimate design should begin with a powerful algorithmic notation to express the recurrence and parallelism associated with the description of the space-time activities. Next, this description will be converted into a optical/VLSI hardware description, or into executable array processor machine codes.

To derive an SFG array for a given algorithm, we can divide the design process into two stages [18]:

1. Deriving a local data dependence graph (DG) from the given algorithm.
2. Mapping the DG onto a signal flow graph (SFG) array by projection.

Figure 1 sketches the two levels in such a design process.

1.1.1 Stage 1: Dependence Graph (DG) Design

A (data) dependence graph (DG) is a directed graph which specifies the data dependencies of an algorithm. In a DG, nodes represent computations, and arcs specify the data dependencies between computations. In our notation, with respect to a dependence arc, the terminating node depends on the initiating node. From the initial sequential algorithm description, a DG for the algorithm can be derived by first converting the algorithm to a *single assignment form*, in which the value of any variable is assigned only once in the algorithm. A single assignment form can clearly show the data dependencies in the algorithm.

For many regular and recursive algorithms we are interested in, the DGs are very likely to be *shift-invariant* or *homogeneous*. A DG is *shift-invariant (SI)* if the dependence arcs corresponding to all the nodes in the index space do not change with respect to the node positions. In this case, the nodes in the DG can be embedded in an index space; therefore, the nodes can be specified by simple indices, such as (i, j, k) . An edge in the DG can be specified by the index difference of the two nodes it connects. For a SI DG, the edges can be simply represented by a small set of *dependence vectors*, each of which specifies the index difference associated with the edge.

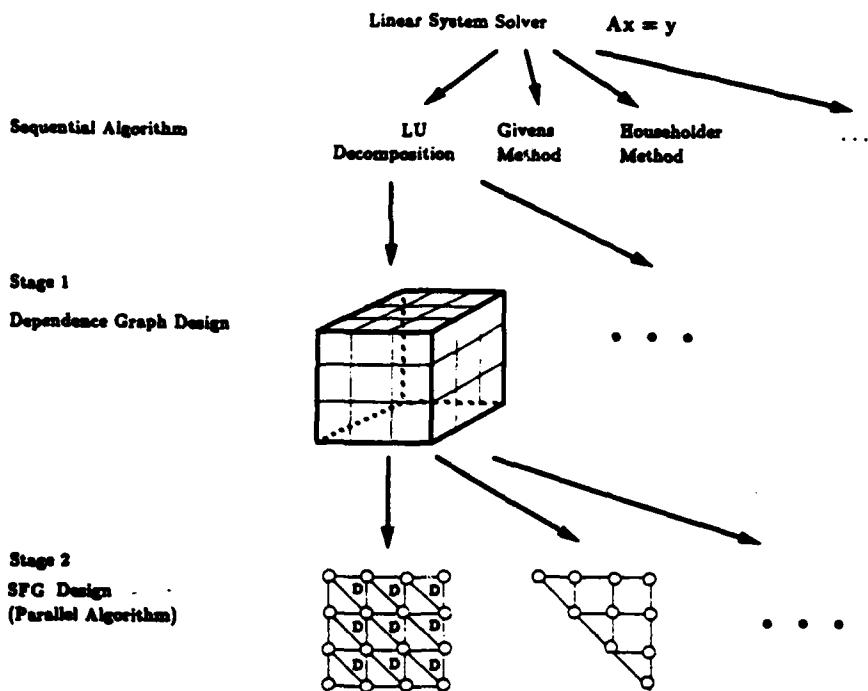


Figure 1: Different levels in designing SFG arrays.

A canonical mapping methodology is used for mapping SI DGs onto processor arrays. Before we can map the SI DG, we first localize the dependence edges, since long communication links are to be avoided in optical/VLSI systems.

Localized DGs

The recursive algorithms, as popularly adopted, are usually not directly given in a localized form. A nonlocalized recursive algorithm, when mapped onto an array processor, is likely to result in an array with global interconnections. Although, in certain instances, such global arcs can be avoided by using a proper projection direction in the mapping schemes. To guarantee a locally interconnected array, a localized recursive algorithm and, equivalently, a localized DG would be derived. Most existing approaches to deriving locally recursive algorithms are rather heuristic. In the following, a more formal (but admittedly still somewhat heuristic) strategy is proposed.

Broadcast Data: If a variable is to be broadcast, then there exists, a corresponding set of index points on which the same data value repeatedly appears. This set is termed a *broadcast contour*. The data are called *broadcast data*. A DG may contain a number of such broadcast contours.

Transmittent Data: The key point is that instead of broadcasting the (public) data along a global arc, the same data may be propagated via local arcs and thus become *transmittent data*. By replacing the broadcast contours by local arcs, a global DG, can be converted to a localized version. This is demonstrated by the following example.

Example: DG for Convolution

The problem of convolution is defined as follows: Given two sequences $u(j)$ and $w(j)$, $j = 0, 1, \dots, N - 1$, the convolution of the two sequences is

$$y(j) = \sum_{k=0}^j u(k)w(j-k)$$

or

$$y_j = \sum_{k=0}^j u_k w_{j-k}$$

where $j = 0, 1, \dots, 2N - 2$

The first step of deriving the recursive equation is to introduce a recursive variable y_j^k . Then the convolution equation can be rewritten in terms of a recursive form:

$$y_j^k = y_j^{k-1} + u_k \cdot w_{j-k} \quad (1)$$

for $j = 0, 1, \dots, N - 1$, and $k = 0, 1, \dots, j$ and for $j = N, N + 1, \dots, 2N - 2$, and $k = j - N + 1, j - N + 2, \dots, N - 1$

Note that Eq. 1 is already in a single assignment form; therefore, the DG can be readily sketched, as shown in Figure 2(a). Equation 1 is an expression with global data dependencies and it is therefore not a locally recursive algorithm.

By replacing the broadcast contours by local arcs, the (global) DG, as shown in Figure 2(a), can be easily converted to a localized version as shown in Figure 2(b), which obviously performs the same algorithm. The localized DG has a corresponding locally recursive single assignment code;

$$y_j^k = y_j^{k-1} + u_j^k \cdot w_j^k, \quad y_j^0 = 0$$

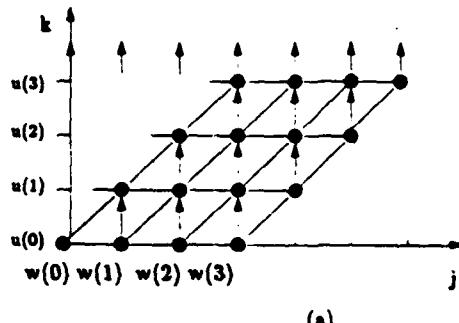
$$u_j^k = u_{j-1}^k, \quad u_0^k = u_k$$

$$w_j^k = w_{j-1}^{k-1}, \quad w_{j-k}^0 = w_{j-k}$$

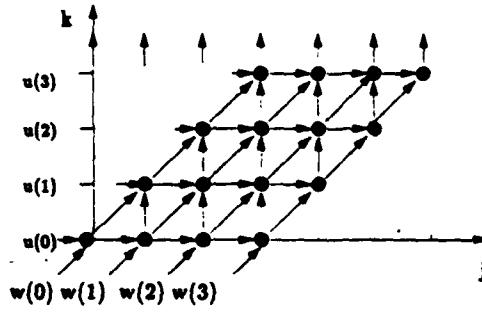
for $j = 0, 1, \dots, N - 1$, and $k = 0, 1, \dots, j$ and for $j = N, N + 1, \dots, 2N - 2$, and $k = j - N + 1, j - N + 2, \dots, N - 1$

1.1.2 Stage 2: SFG Array Design

To determine a valid array structure for a SI local DG, one straightforward design method is to designate one processing element (PE) for each node in a DG. This however in general leads to very inefficient utilization of the PEs, since each PE can be active only for a small fraction of the computation time. In order to improve PE utilization, it is often desirable



(a)



(b)

Figure 2: DGs for convolution: (a) global; (b) localized.

to map the nodes of the DG onto a fewer number of PEs. To achieve this it is useful to map the DG first to an intermediate expression, i.e., in a SFG form.

SFG Arrays

The SFG expression, which consists of processing nodes, communicating edges, and delays. In general, a *node* is often denoted by a circle representing an arithmetic or logic function performed with zero delay, such as multiply or, add. An *edge*, on the other hand, denotes either a dependence relation or a delay. Unless otherwise specified, the following conventions are adopted for convenience. When an edge is labeled with a capital letter D (or D' , $2D$, ...), it represents a time delay operator with delay time D (or D' , $2D$, ...). The SFG representation derives its power from the fact that the computations are assumed to be delay-free, i.e., they take no time at all. Consequently, the burden of tracing the detailed time-space activities associated with pipelining are eliminated. Moreover, any delay in the system has to be explicitly introduced in the form of so-called delay branches. These delay branches allow history-sensitive systems to be described in a clear and unambiguous way.

Projection and Linear Scheduling

There are two important considerations needed to be taken into account in the mapping from a DG to an SFG. One is the *node assignment*, and the other is the *scheduling*. The node assignment involves specifying the task to be performed in one cluster of nodes in DG be assigned to a single node in the SFG. The scheduling involves specifying the sequence of operations for all the nodes in the DG.

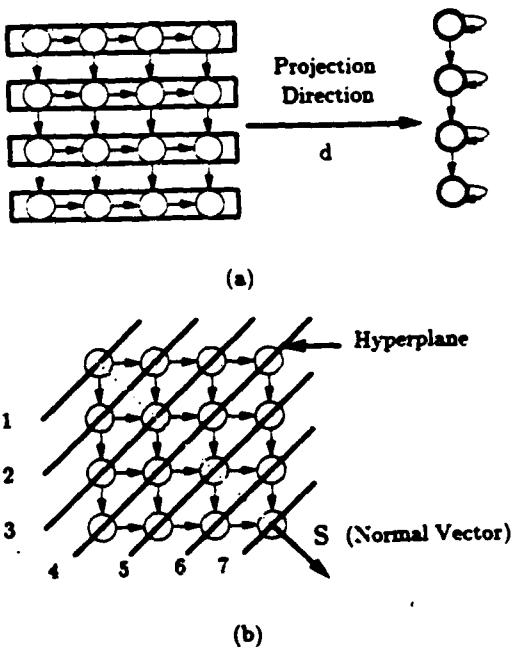


Figure 3: Projection with projection vector \vec{d} ; (b) a linear schedule \vec{s} and its hyperplanes.

For a SI DG, we can use linear node assignment (projection) and linear scheduling for the mapping, both of which will preserve the regular structure of the DG.

Projection A linear assignment (projection) of a DG is a linear mapping from the nodes of the DG to nodes in the SFG, in which nodes along a straight line in the DG are mapped to a single node in the SFG. The projection direction is represented by the projection vector \vec{d} . Mathematically, a projection in a direction \vec{d} is represented by an $n \times (n - 1)$ projection matrix P , which is orthogonal to the projection vector \vec{d} , i.e., $P^T \vec{d} = \vec{0}$. The P matrix maps a n dimensional node index vector i to an $n - 1$ dimensional vector $i' = P^T i$ (see Figure 3(a)).

Linear Scheduling In a schedule, the scheduled execution time of a node is represented by a time index (i.e. an integer). The schedule is a linear schedule if a set of parallel equi-temporal hyperplanes, specified by the normal vector \vec{s} , are mapped to a set of linearly increased time indices, then the time index of the nodes can be mathematically represented by $s^T i$ (see Figure 3(b)).

Permissible Linear Schedules

Given a DG and a projection direction \vec{d} , we note that not all hyperplanes are qualified to define a valid schedule for the DG. Some of them violate the precedence relation of

computation specified by the dependence arcs. The allowable directions of the hyperplanes actually define the class of *permissible linear schedules*. In order for the given hyperplanes to represent a permissible linear schedule, it is necessary and sufficient that the normal vector \vec{s} satisfies the following two conditions:

$$(1) \quad \vec{s}^T \vec{e} \geq 0, \text{ for any dependence arc } \vec{e}. \quad (2)$$

$$(2) \quad \vec{s}^T \vec{d} > 0. \quad (3)$$

Both the conditions Eq. 2 and Eq. 3 can be checked by inspection. In short, the *schedule is permissible if and only if* (1) *all the dependency arcs flow in the same direction across the hyperplanes; and* (2) *the hyperplanes are not parallel with projection vector \vec{d} .* The first condition means that a causality should be enforced in a permissible schedule. Namely, if node p depends on node q, then the time step assigned for p cannot be less than the time step assigned for q. The second condition implies that nodes on an equitemporal hyperplane should *not* be projected to the same PE. The permissible hyperplane directions defined by the first condition is the same as the notion of the time cone in [7] and [3].

Example: SFG for Convolution

An SFG for convolution, by projecting the DG onto the $\vec{d}^T = [1, 1]$ direction and with scheduling vector $s = [0, 1]$, is shown in Figure 4. Note that the existence of the delay-free edges means that the data will be propagated in "zero time". The computation in each node is also assumed to be done in "zero time".

2 Applications to Image Feature Extraction

In this section, we demonstrate this top-down design methodology by systematically mapping the applicational algorithms onto appropriate VLSI arrays. The applications we are interested is the image and video signal processing. More specifically, the following algorithms are treated in this section.

- Edge detection via 2-D/1-D convolution.
- Video enhancement via median/rank order filtering.

2.1 Edge Detection for Feature Extraction

Most of the useful information in an image is contained in those regions where a change of gray levels or colors occurs, i.e., at the edges. The information consists of the size of the transition (edge magnitude) and the direction in which the intensity changes most rapidly (edge direction). These quantities can be computed from the partial derivatives of image function $g(x, y)$. The information obtained from applying the derivative operators allows segmentation of an image, which is an important step toward object classification and identification. The extracted edge information can be used to determine the contour

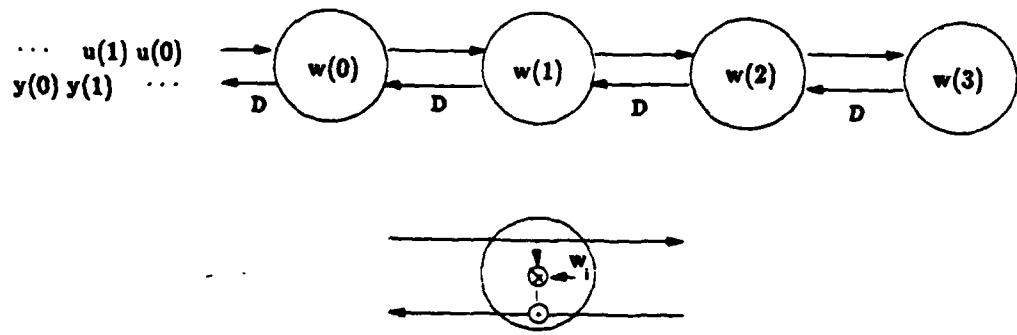


Figure 4: An SFG for convolution.

of objects, and under certain conditions, the topological features of 3-D objects can be determined. Edge information can also be used to evaluate threshold values to isolate specific image regions.

2.1.1 Edge Detection by means of 2-D Convolution

Two-dimensional convolution is the most common way for image edge detection. Assume that an image g is to be edge detected, the 2-D convolution involves convolving a small pattern (mask) f with the larger image g . This is done by moving the window pattern f over the image, as shown in Figure 5. At each point the convolution is computed:

$$c(m, n) = \sum_i \sum_j f(i, j)g(m - i, n - j) \quad (4)$$

Various local operators have been widely used, such as Roberts operators, Laplacian operators, Prewitt operators, and the Sobel operators (see Figure 6). The Laplacian operators detect the edges as the local maxima of the image gradients. Therefore, zero-crossing of the second derivatives gives the edge points. Furthermore, the Laplacian operator can pick up edges from all possible directions (isotropic) by a simple convolution with the image and potentially produce very thin edges.

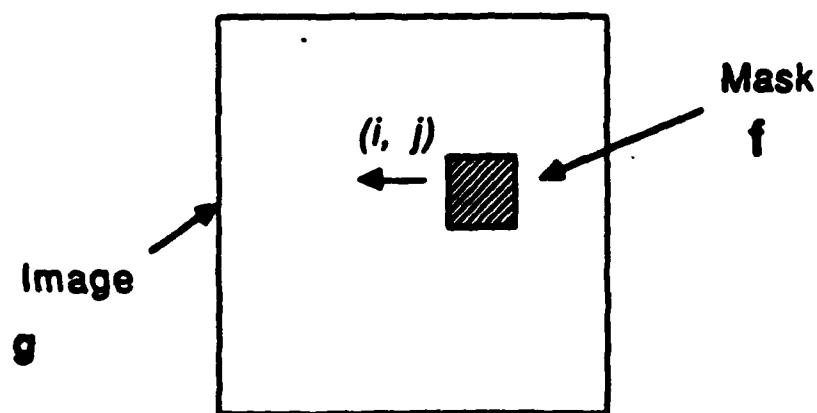


Figure 5: Moving windows for image-edge detection.

$$\begin{aligned}
 W_{PV} &= \begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix} \\
 W_{PV} &= \begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix} \\
 W_{SV} &= \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \\
 W_{SV} &= \begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix} \\
 W_L &= \begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}
 \end{aligned}$$

(a) (b) (c) (d)

Figure 6: Various windows used in edge detection: (a) Prewitt operators; (b) Sobel operators; (c) Laplacian operator; (d) expanded Laplacian operators.

2.1.2 Array Design for 2-D Convolution

For 2-D convolution we define a $k_1 \times k_2$ matrix as a window. Since the size of an image may typically vary from 256×256 up to $8K \times 8K$ pixels, we immediately see the necessity of massive parallel processing. For real time applications, where the image is updated 30 to 60 times per second, the speed requirements are even more critical. In the following, an SFG design for the 2-D convolution is described.

DG Design Without loss of generality, assume that the window size is 3×3 and the image size is $N \times N$. Then Eq. 4 can be rewritten as:

$$c(m, n) = \sum_{i=0}^2 \sum_{j=0}^2 f(i, j)g(m - i, n - j) \quad (5)$$

Since there are four indices, i, j, m, n , in this equation, the DG for this algorithm is 4-D (four dimensional) and the SFGs obtained via projection are 3-D. Because of the difficulty in displaying the 4-D DG, we first decompose the double summation term in Eq. 5 into three terms

$$\begin{aligned} c(m, n) &= \sum_{j=0}^2 f(0, j)g(m, n - j) + \sum_{j=0}^2 f(1, j)g(m - 1, n - j) \\ &\quad + \sum_{j=0}^2 f(2, j)g(m - 2, n - j) \end{aligned} \quad (6)$$

each summation term in Eq. 6 may be described by the 3-D DG as shown in Figure 7(a). The 3-D DG represents the image convolving with one row of the window pattern (1×3). Note that each 2-D ($3 \times N$) layer of the 3-D DG is basically a DG for 1-D convolution. The 4-D DG may be constructed by summing up the outputs of three such similar 3-D DGs.

SFG Design Let us first consider the design of a 3-D SFG with (approximate) size $3 \times 3 \times N$. This 3-D SFG may be obtained by first projecting each of the three 3-D DGs to one 2-D SFG (of size $3 \times N$) (see Figure 7(b)) and then combining these three 2-D SFGs by summing up their outputs (see Figure 7(c)). If 2-D SFGs are preferred, then multiprojection can be adopted to project the 3-D SFG to 2-D SFGs. Figure 7(d) shows a 2-D SFG with size $3 \times N$. Different projection directions can also be applied if desired. For example, it is apparent that a 3×3 SFG can also be obtained by changing the direction of multi-projection.

2.1.3 Use 1-D Arrays to Perform 2-D Edge Detection

In general the classic gradient or Laplacian operators perform poorly on noisy images. To improve the performance, one solution is to do averaging before the application of

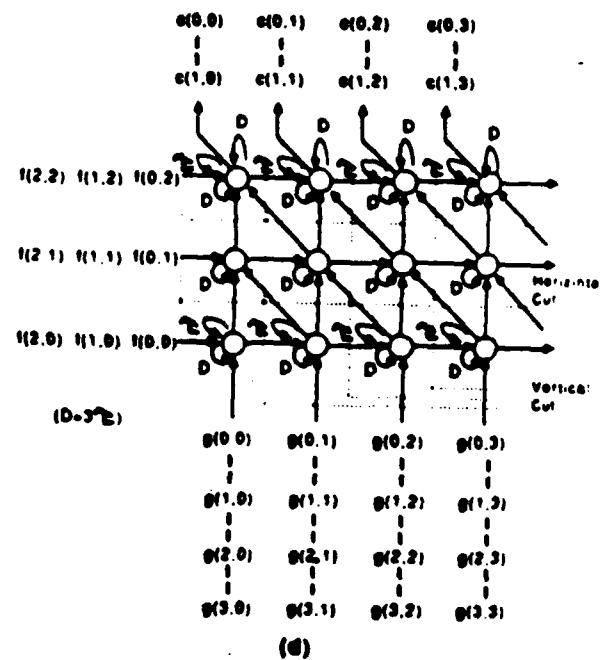
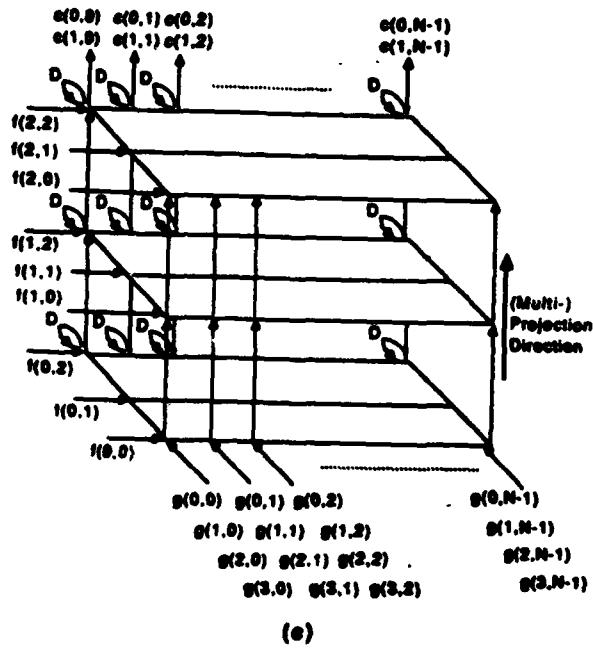
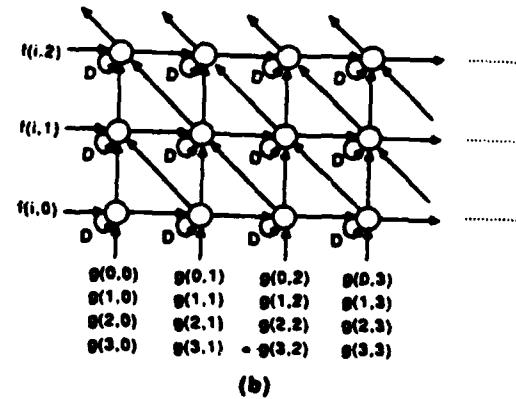
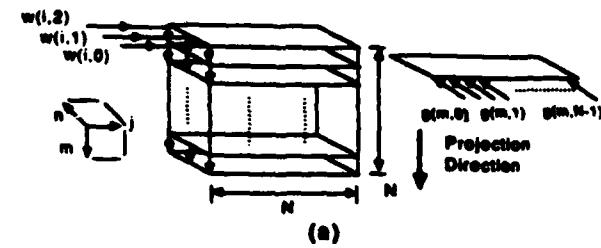


Figure 7: (a) A 3-D DG for the image ($N \times N$) convolved with one row of the window pattern (1×3). The 4-D DG may be constructed by summing up the outputs of 3 such 3-D DGs; (b) A 2-D SFG obtained by projecting the 3-D DG along m -direction; (c) The 3-D SPG obtained by summing up the outputs of 3 2-D SFGs; and (d) The 2-D SPG obtained by adopting multiprojection.

the operators. This results in an expanded Laplacian operator with 9×9 mask size (see Figure 6(d)), which can be used in noisy averaging and edge detecting. Marr and Hildreth also suggest a family of edge operators based on the zero-crossing of a generalized Laplacian [24], which are also isotropic second derivative zero-crossing operators. Figure 8(a) shows an example of an 11×11 generalized Laplacian operator. When a larger mask operator is adopted, the performance is in general improved but at the expense of higher hardware cost. To overcome this hardware problem, an alternative is proposed below.

Note that most of the large mask operators share a low rank property. For example, the expanded 9×9 Laplacian operator shown in Figure 6(d) has only rank two. The generalized 11×11 Laplacian operator shown in Figure 8(a) has numeric rank two, based on the distribution of their singular values as shown in Figure 8(b). (Note that the first two singular values are significantly larger than the rest.)

Let p denote the numeric rank of a $k \times k$ mask operator W , e.g., $k = 9$ or 11 and $p = 2$ in the above examples. Then W can be approximated by a linear combination of p outer products, (see Eq. 2.17)

$$W \approx \sum_{i=1}^p \sigma_i u_i v_i^T = \sum_{i=1}^p \hat{u}_i \hat{v}_i^T \quad (7)$$

where u_i and v_i are the singular vectors of W , and \hat{u}_i and \hat{v}_i are the singular vectors scaled by the square root of the singular value σ_i . For example, the scaled singular vectors of the expanded 9×9 Laplacian operator are

$$\begin{aligned}\hat{u}_1 &= [1 \ 1 \ 1 \ -4 \ -4 \ -4 \ 1 \ 1 \ 1] \\ \hat{v}_1 &= [0 \ 0 \ 0 \ -1 \ -1 \ -1 \ 0 \ 0 \ 0] \\ \hat{u}_2 &= [0 \ 0 \ 0 \ -1 \ -1 \ -1 \ 0 \ 0 \ 0] \\ \hat{v}_2 &= [1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1]\end{aligned}$$

Eq. 7 means that the 2-D edge detection can be implemented by a combination of cascaded 1-D column convolutions and 1-D row convolutions. More elaborately, the image is first processed by a row convolution then followed by a column convolution. Mathematically, let $W * g$ denote the 2-D convolution of the mask operator W and the image g , then based on Eq. 7

$$W * g \approx \hat{u}_1 *_c (\hat{v}_1 *_r g) + \hat{u}_2 *_c (\hat{v}_2 *_r g)$$

where $*_c$ and $*_r$ denote 1-D column convolution and 1-D row convolution respectively.

The array processor design for the cascaded row/column convolutions may be easily carried out by using the canonical mapping methodology. Since $p \ll k$ in general, this approach can yield a significant hardware saving compared with the 2-D convolution design.

-34.	-25.	-16.	-12.	-8.	-7.	-6.	-12.	-16.	-25.	-34.
-25.	-15.	-7.	-1.	3.	4.	3.	-1.	-7.	-15.	-25.
-16.	-7.	2.	8.	12.	14.	12.	8.	2.	-7.	-16.
-12.	-1.	0.	15.	20.	21.	20.	15.	0.	-1.	-12.
-8.	3.	12.	20.	24.	26.	24.	20.	12.	3.	-8.
-7.	4.	16.	21.	26.	27.	26.	21.	16.	6.	-7.
-6.	3.	12.	20.	24.	26.	24.	20.	12.	3.	-6.
-12.	-1.	8.	15.	20.	21.	20.	15.	8.	-1.	-12.
-16.	-7.	2.	8.	12.	14.	12.	8.	2.	-7.	-16.
-25.	-15.	-7.	-1.	3.	4.	3.	-1.	-7.	-15.	-25.
-34.	-25.	-16.	-12.	-8.	-7.	-6.	-12.	-16.	-25.	-34.

(a)

126.72 116.24 2.00 1.39 0.50 0.41 0.00 0.00 0.00 0.00 0.00

(b)

-34.18-24.09-17.70-12.05 -6.13 -7.10 -6.13-12.05-17.70-24.09-34.18
-24.09-15.05 -7.10 -1.04 3.04 6.24 3.04 -1.04 -7.10-15.05-24.09
-17.70 -7.10 1.04 8.12 12.02 13.00 12.02 8.12 1.04 -7.10-17.70
-12.05 -1.04 8.12 15.21 19.66 21.17 19.66 15.21 8.12 -1.04-12.05
-6.13 3.04 12.02 19.66 24.15 25.72 24.15 19.66 12.02 3.04 -6.13
-7.10 4.24 13.00 21.17 25.72 27.33 25.72 21.17 13.00 4.24 -7.10
-3.13 3.04 12.02 19.66 24.15 25.72 24.15 19.66 12.02 3.04 -3.13
-12.05 -1.04 8.12 15.21 19.66 21.17 19.66 15.21 8.12 -1.04-12.05
-17.70 -7.10 1.04 8.12 12.02 13.00 12.02 8.12 1.04 -7.10-17.70
-24.09-15.05 -7.10 -1.04 3.04 6.24 3.04 -1.04 -7.10-15.05-24.09
-34.18-24.09-17.70-12.05 -6.13 -7.10 -6.13-12.05-17.70-24.09-34.18

(c)

Figure 8: The singular value decomposition of generalized Laplacian operator: (a) the original operator; (b) the singular values distribution; (c) the reconstructed operator using the first two ranks only.

2.2 Median/Rank-Order Filtering for Image Enhancement

2.2.1 Median Filter for Video Signal Processing

Median filters are a special class of nonlinear filters useful in the removal of impulse noise from signals [27]. The idea of the median filter is very simple. Take a sampled signal of length L and slide it across a window of length $2N + 1$, the filter output at each window position is given the same position as the center sample of the window and is set equal to the median value of the $2N + 1$ samples in the window. Start-up and end effects are taken care of by appending samples of constant value to the beginning and the end of the data sequence. As an example consider the sequence

3 4 7 1 2

The median is found by first sorting this sequence and then choosing the center value of the sorted sequence. Hence the original center pixel value of the processing window (here of length five), "7", is replaced by the median of the sequence, "3". An attractive property of median filters for image processing is that it removes sparkle noise, while preserving edges. This means that while the traditional linear filtering is unable to preserve the scene changes, the median filtering has no such problems. Consider the example shown in Figure 9(a), where the scene change may be perceived as an edge, a median filter of window length 3 leads to a result shown in Figure 9(b). This property may be exploited in frame to frame basis video signal processing, where scene changes need to be preserved.

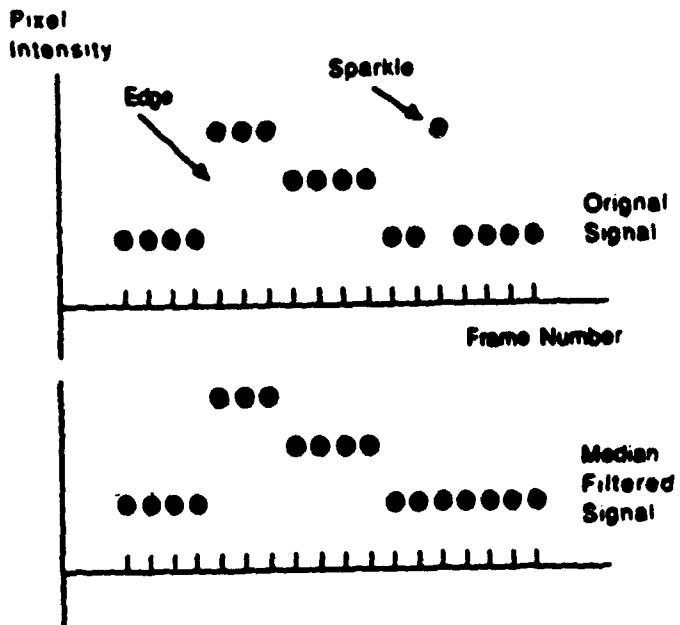


Figure 9: (a) Video signal with three level scene change on a frame to frame basis.
 (b) Median filtered video signal (with filter window length 3).

2.2.2 Noise Removal via Median/Rank Order Filtering

One approach to the median filtering is to use an SFG sorting arrays and to sort out the median value of a data window [4], [25]. The window then slides one pixel further to let the sorter generate the next median value and so on. The bubble-sorter design [18] can be adapted to realize such moving window characteristics for the median filtering. This kind of approach can be classified as being *window-oriented*, in which the pixels in every window are separately sorted. Most window-oriented designs actually use 2-D sorting arrays. It is wasteful and is not appealing to large windows, because of the large array size and low utilization rate.

Rank Order Filtering A more efficient *pixel-oriented* approach is rank order filtering, where the ranks of the intensities of the pixels are determined in descending order. The pixel that has the required rank is extracted as output, while the ranks of all but one of the pixels in the window are useful for the rank-ordering of the next window.¹

Consider a sequence of pixels and a 1-D window W of length k . As the window slides into the next one, all the pixels except the departing one appear again in the next window. Therefore, if the ranking in the present window is evaluated and stored, the ranking in the next window can be determined by a simple modification of the present ranking rather than

¹This approach is applicable for extracting not only the median value but also data of any specified rank.

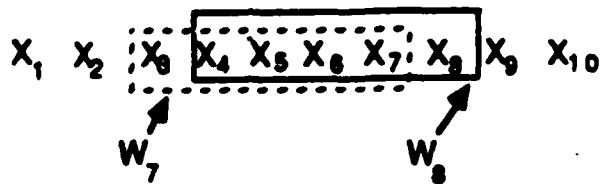


Figure 10: A window of length 5 for rank ordering and its next window.

a reevaluation from scratch. With reference to Figure 10, suppose that the window length is $k = 5$ and the present window considered is $W_7 = [x_3 \ x_4 \ x_5 \ x_6 \ x_7]$. The next window $W_8 = [x_4 \ x_5 \ x_6 \ x_7 \ x_8]$ is obtained from W_7 by deleting the pixel x_3 and inserting the pixel x_8 . Suppose that the ranks of pixels $\{x_4, x_5, x_6, x_7\}$ are previously evaluated in W_7 , say, as $\{r_{i7}, i = 4, 5, 6, 7\}$. Therefore, their new ranks $\{r_{i8}\}$ in the window W_8 can be determined by comparing the pixels $\{x_i, i = 4, 5, 6, 7\}$ with the departing pixel x_3 and the new arrival x_8 , and accordingly modifying $\{r_{i7}\}$ into $\{r_{i8}\}$. This is called the PDI (pixel deletion insertion) process. In the same process, the rank of pixel x_8, r_{88} , can also be obtained. Each window requires $O(k)$ comparisons for the pixel-oriented approach, compared with $O(k^2)$ comparisons for the window-oriented approach.

2.2.3 Array Design of a Rank Order Filter

Our discussion will be based on one example showing that the ranks in one window, e.g., $\{r_{i7}\}$, lead to the ranks in the next window $\{r_{i8}\}$. Then by induction, the overall procedure should follow. Suppose that after the 7-th recursion, the ranks $\{r_{37}, r_{47}, \dots, r_{77}\}$ are available² and the data are arranged in a reversed order

$$z_7 \ z_6 \ z_5 \ z_4 \ z_3$$

Now the new input data z_8, z_9, \dots are available from the right side. Then the DG for the PDI procedure is shown in Figure 11(a), and the corresponding SFG obtained by projecting the DG in the vertical direction is shown in Figure 11(b). As shown in the DG, the PDI procedure consists of two processing phases (phases A and B). For the following illustration of the processing in the SFG, we note that, for the 8-th recursion, z_3 is the *deletion pixel* and z_8 is the *insertion pixel*.

²Start-up can be taken care of by appending constant samples (e.g., 0's) to the beginning of the sequence.

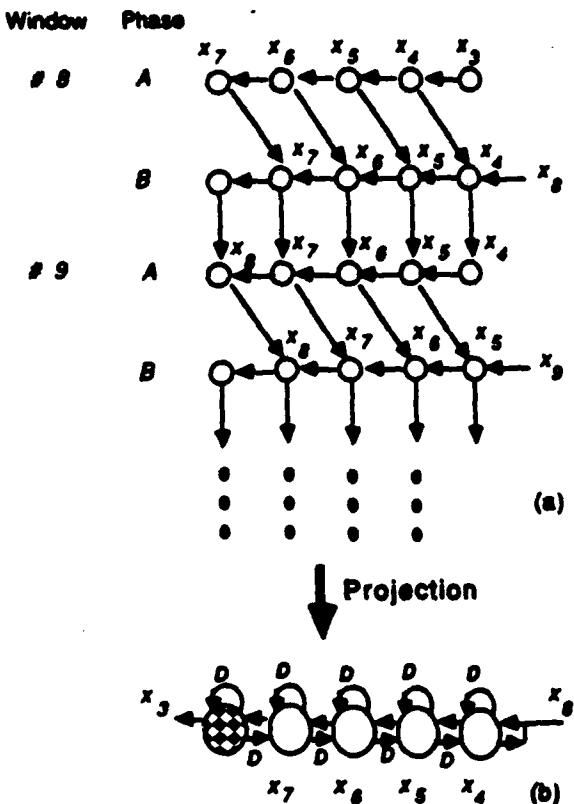


Figure 11: (a) DG for the pixel deletion/insertion procedure. (b) The corresponding SFG obtained by projecting DG in the vertical direction.

Pixel Deletion Rank Updating In phase A of the 8-th recursion, the data x_3 is propagated to the left, meeting all the other data $\{x_4, x_5, x_6, x_7\}$ in their respective PEs. The PEs perform the following function:

$$\begin{aligned}
 \text{For } i &= 4, 5, 6, 7 \\
 \text{update } r_{i7} \\
 r'_{i7} &= r_{i7} - 1 && \text{if } x_i \geq x_3 \\
 r_{i7} &= r_{i7} && \text{otherwise}
 \end{aligned}$$

We note that $\{x_i, r'_{i7}\}$ should be moved to the right neighbor PE after each update to be ready for the phase B processing.

Pixel Insertion Rank Updating In the phase B of the 8-th recursion, the data x_8 is propagated to left meeting all the other data $\{x_4, x_5, x_6, x_7\}$ in the array in their respective PEs. The PEs perform the following function:

For $i = 4, 5, 6, 7$

update r_{is}

$$\begin{aligned} r_{is} &= r'_{i7} + 1 && \text{if } x_i < x_8 \\ r_{i7} &= r_{i7} && \text{otherwise} \end{aligned}$$

Rank Evaluation of New Pixel Note that x_8 is the new arriving pixel, and its rank must be evaluated. This evaluation process can be accomplished while the pixel insertion process is taking place. The procedure is described as follows:

Set $t_{s3} = 1$ initially
For $i = 4, 5, 6, 7$

update t_{si}

$$\begin{aligned} t_{si} &= t_{s,i-1} && \text{if } x_i < x_8 \\ t_{si} &= t_{s,i-1} + 1 && \text{otherwise} \end{aligned}$$

propagate t_{s3} to left
Output $r_{s3} = t_{s7}$

Here, an auxiliary variable t_{si} is introduced for the purpose of counting the number of pixels greater than or equal to x_8 . After all the comparisons are performed, the rank for x_8 can be determined, $r_{s3} = t_{s7}$. We note that it is easy to combine the rank evaluation process with the pixel insertion process.

Rank Extraction of the Current Window The rank extraction problem here is to determine the integer i^* , such that r_{i^*s} equals the desired rank and extract x_{i^*} from the array. For convenience of pipeline array design, it is proposed that the data x_{i^*} , once identified, should be propagated leftward and eventually output from the leftmost boundary PE. In our design, the PDI, rank evaluation, and rank extraction processes are all embedded into one linear array.

2.2.4 2-D Median Filtering

Image enhancement for video signal processing applications can make good use of 1-D median filtering performed on a frame to frame basis. However, for most other image processing applications, we still require filtering to take place in two dimensions. The key properties of the 2-D median filter are mostly preserved by using two separated 1-D median filters, one filter processes pixels along the rows and then the resulting image is processed by the other filter along the columns. Thus 2-D median filtering can virtually be performed by using two 1-D SFG arrays [26] [28].

3 Array Design for Artificial Neural Network

Neuroscientists have revealed that the massive parallel processing power in the human brain lies in the global and dense interconnections among a large number of identical logic elements or *neurons*. These neurons are connected to each other with variable strengths by a network of *synapses* [33]. The original discrete-state Hopfield model [29], and the continuous-state Hopfield-Tank model (including the proposed analog neural circuits) [30],[31] have recently become popular in the realization of artificial neural networks (ANNs). They can be programmed to perform computational networks for associative retrieval or for optimization problems. Several optical computing approaches, exploiting the global interconnectivity of optical signal flow, have been proposed [34], [35], [36].

Both the analog neural circuits and the optical neural networks suffer from the disadvantages of low precision, difficulty of modifying the synaptic strengths, convergence to local optima and global interconnectivity. In order to overcome the disadvantages mentioned above, a locally interconnected SFG array for an ANN is proposed [11].

3.1 Hopfield and Hopfield-Tank Models

Original Hopfield Model The original Hopfield model used two-state threshold *neurons*. (The states are usually taken to be 0 and 1). In this discrete model, each neuron i receives input currents of $T_{ij}V_j$ from neuron j , and a bias current input I_i (see Figure 12(a)) [29], [33]. The discrete-time transition of neuron i can thus be formulated as:

$$\begin{aligned} U_i(k) &= \sum_j^N T_{ij}V_j(k) + I_i \\ V_i(k+1) &= \text{step}[U_i(k)] \end{aligned} \quad (8)$$

where $\text{step}[x]$ is a unit step function, which is 1 for $x \geq 0$ and 0 for $x < 0$. N is the number of neurons and T_{ij} are elements of an interconnection matrix representing the synaptic strengths of neural connections (with $T_{ii} = 0$). The inputs U_i are thresholded by the amplifier (neuron), and the output V_i can then be fed back to the input of any other amplifier to modify their states. The state space transitions will gradually converge to a set of stable fixed points.

Hopfield and Tank Model In order to imitate the continuous input-output relationship of real neurons, and also to simulate the integrative time delay due to the capacitance of real neuron, the continuous-state model proposed by Hopfield and Tank can be approximated by the following dynamic equations [30], [31], [33]:

$$U_i(k) - U_i(k-1) = \sum_j^N T_{ij}V_j(k) + I_i \quad (9)$$

$$V_i(k+1) = g[U_i(k)] \quad (10)$$

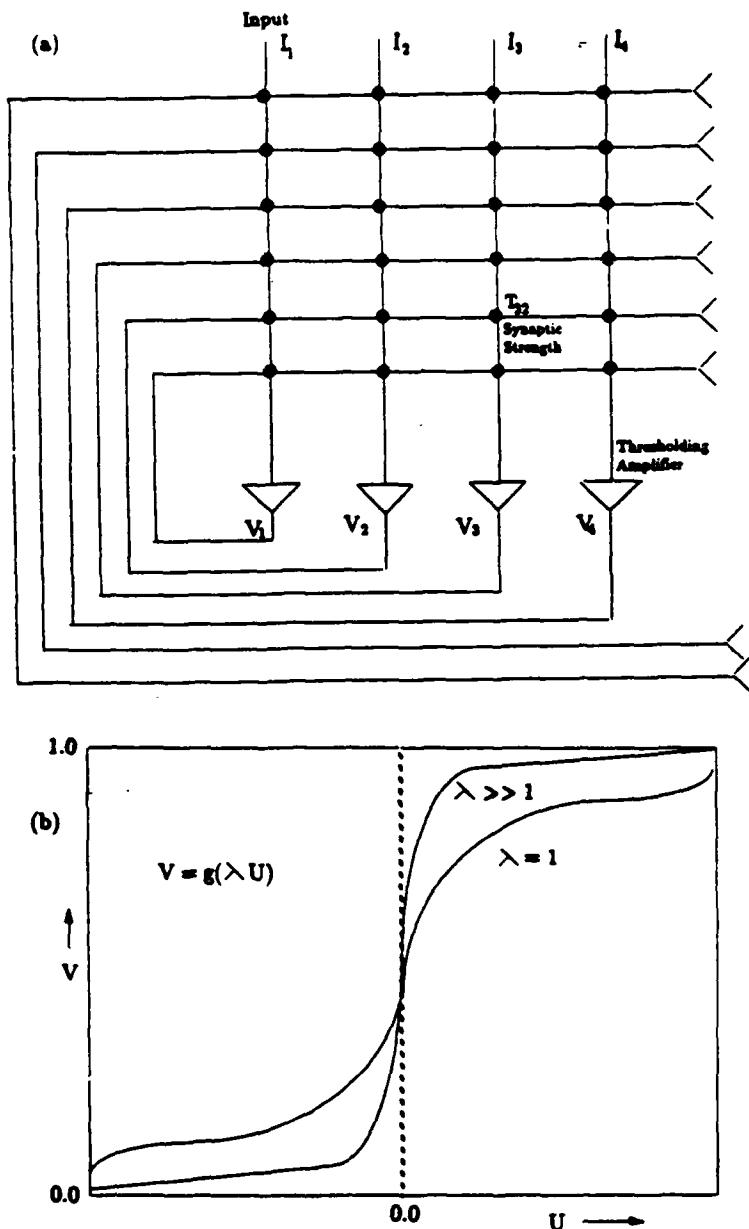


Figure 12: (a) A simplified schematic diagram of the Hopfield model. In the discrete-state model, the amplifiers act like a hard limiter, while in the continuous-state model, the amplifiers follow the input-output relationship $g_i[\lambda u]$ with dynamic behavior. (b) The typical input-output relationship, the λ value determines the gain of the thresholding function.

where $g[x]$ is a nonlinear function, e.g.,

$$g[x] = (1/2)[1 + \tanh(x/x_0)] \quad (11)$$

which approaches a unit step function as x_0 tends to zero (see Figure 12(b)). The right hand side of Eq. 9 can be considered as the new excitation source, which effects "modification" of the states as shown in the left hand side.

For both models, an energy function is defined as [29], [30]

$$E = -(1/2) \sum_{i=1}^N \sum_{j=1}^N T_{ij} V_i V_j - \sum_{i=1}^N I_i V_i \quad (12)$$

Hopfield has shown that, if $T_{ij} = T_{ji}$, then neurons in the continuous-state model always change their states in such a manner that the energy function is reduced [30]:

Matrix Formulation- The Hopfield model can be formulated as a consecutive matrix-vector multiplication problem with some prespecified thresholding operations [34]. For example, a matrix-form expression of Eq. 9 and Eq. 10 can be written as

$$\begin{aligned} u(k) &= \mathbf{T}v(k) + \mathbf{i} + u(k-1) \\ v(k+1) &= G[u(k)] \end{aligned} \quad (13)$$

where $G[\mathbf{x}]$ function specifies the nonlinear thresholding of each element of the vector \mathbf{x} , and the vectors and matrices used are given as:

$$\begin{aligned} \mathbf{u} &= [U_1, U_2, \dots, U_N]^T \\ \mathbf{v} &= [V_1, V_2, \dots, V_N]^T \\ \mathbf{i} &= [I_1, I_2, \dots, I_N]^T \\ \mathbf{T} &= \left[\begin{array}{cccc} T_{11} & T_{12} & \cdots & T_{1N} \\ T_{21} & T_{22} & \cdots & T_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N1} & T_{N2} & \cdots & T_{NN} \end{array} \right] \end{aligned} \quad (14)$$

3.2 SFG Design via Cascade DG

Cascade DG Design The consecutive matrix-vector multiplication array architecture design can be derived from a cascaded DG with nonlinear assignment. The DG of Eq. 13 is shown in Figure 13. Using a nonlinear assignment, a locally interconnected SFG array with bidirectional communication links can be obtained. This SFG architecture requires some smart switches to change the operations of each PE at different time slots. However, some T_{ij} 's need to be repetitively stored in multiple PEs, and making the design unsuitable for adaptive modification of T_{ij} .

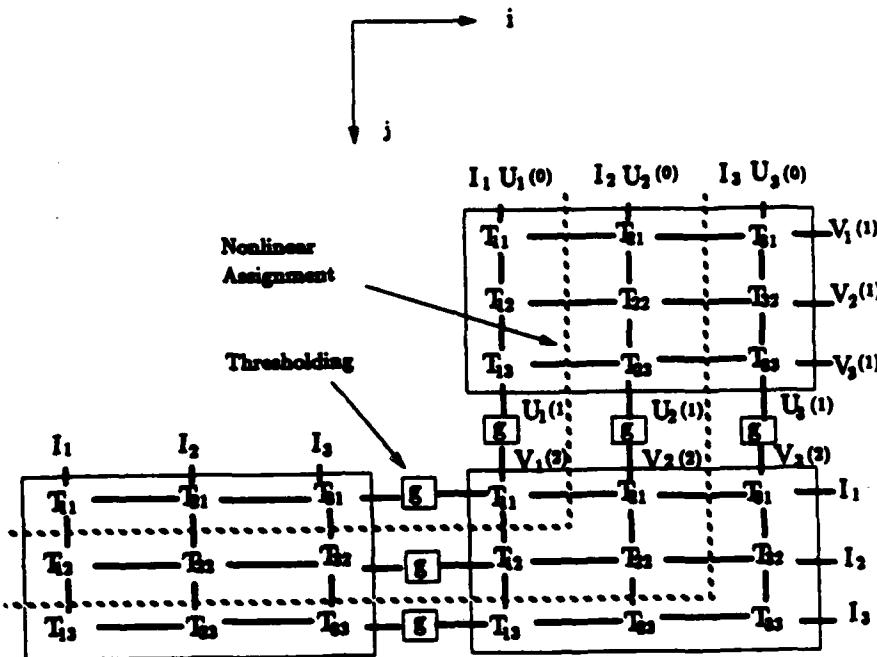


Figure 13: DG for consecutive matrix-vector multiplication formulation of Hopfield model.

Modified Cascade DG Design Examining the DG design shown in Figure 13 more closely, we note that many undesirable design aspects are due to the fact that the input direction of neuron states $V_i(k)$ is orthogonal to the output direction of the thresholded neuron states $V_i(k+1)$. It is possible to rearrange the data ordering of the T_{ij} elements, so that the input direction $V_i(k)$ becomes parallel to the output direction of $V_i(k+1)$. Such a modified DG is depicted in Figure 14. In this DG, for $i = 1, 2, \dots, N$, the i -th column of the T_{ij} data array is circularly shifted-up by $i-1$ positions. This DG is not totally localized due to the presence of the global spiral communication arcs. However, the input direction (from the top) and the output direction (from the bottom) are parallel. The advantage is that when many such DGs are cascaded top-down, the inputs and outputs data can be matched perfectly.

Ring Array for ANN For the top-down cascaded DG, the projection can be taken along the vertical direction, which will result in a ring SFG array as shown in Figure 15.³ In the ANN implementation each SFG node, say the i -th node, is treated as a neuron, and the synaptic strengths ($T_{i1}, T_{i2}, \dots, T_{iN}$) are stored in it. At the k -th iteration, the operation of the node is as follows:

³The same architecture was previously proposed by Porter [38], for consecutive matrix-vector multiplication in solving discrete state space equations.

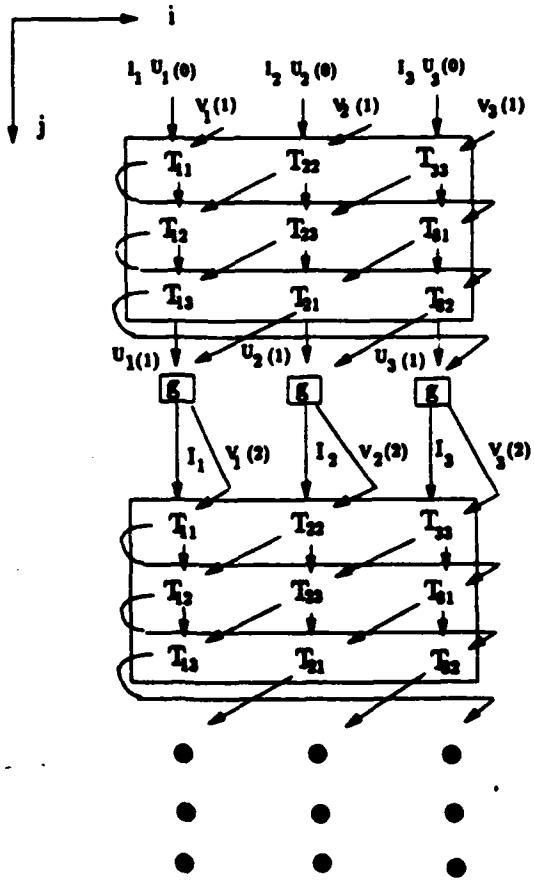


Figure 14: Modified DG for consecutive matrix-vector multiplication formulation of Hopfield model.

1. Each of the neuron outputs (V_1, V_2, \dots, V_N) is cycling through the ring array, and will pass through the i -th node through broadcasting.
2. When V_j passes through the i -th node, it is multiplied with T_{ij} , and the result is added to the sum of $U_i(k - 1)$ and I_i (according to the Eq. 9).
3. After V_j passes through all the nodes, the computation for $U_i(k)$ is completed, and it is ready for the thresholding operation.
4. After the thresholding operations, the PE sends the thresholded neuron output $V_i(k + 1)$ to the left-side neighbor PE.

The above procedures is repeated until a convergence is reached. For implementing a large number of neurons, the problem of the long wrap-around line can be solved by a special 2-D arrangement scheme [11].

Advantageous Properties of SFG ANN The advantages of this array architecture for ANN are summarized as follows:

- Only N synaptic strengths of T_{ij} are stored in each node. This results in easier modification of the synaptic strengths, making the "learning" capability possible.

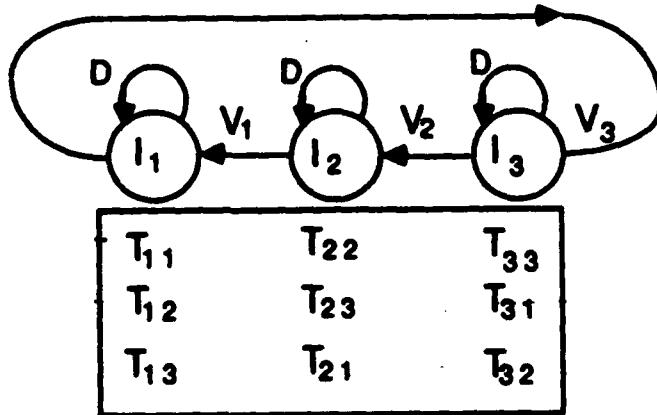


Figure 15: Ring SFG array for consecutive matrix-vector multiplication formulation of Hopfield model.

- Using this proposed SFG architecture, the gain parameters can be easily updated during the iterations. (It is observed that if the gain control parameter of the sigmoid input-output relation in the neuron can be dynamically changed, then a faster convergence and better performance can be obtained [31].)
- Compared to the analog RC neural circuit and the other optical neural networks, the digital implementation can achieve higher precision in computation.

4.3 Applications of ANN in Image Processing and Analysis

Solving Image Restoration Problems using an ANN Artificial neural networks have been successfully applied to low level vision processing [37]. Here, an image restoration example is given to illustrate how to map applicational problems onto ANNs. Consider an observed degraded image vector \mathbf{g} , which can be mathematically expressed as

$$\mathbf{g} = \mathbf{Hf} + \mathbf{n} \quad (15)$$

where \mathbf{H} is a known blurring matrix. The statistical properties of the noise vector \mathbf{n} are also assumed known. Image restoration is the scheme whereby an image vector \mathbf{f} is restored from the linear blurring degradation mechanism and the additive noise. The most straightforward

method is to solve the least squares problem to find an estimate \hat{f} which minimizes the total estimation error.

$$\min (g - H\hat{f})^T(g - H\hat{f}) \quad (16)$$

In many cases, however a priori information about the image properties (e.g., smoothness, intensity distribution) is known. In order that the estimated solution also reflects this information, a modified least squares formulation should be adopted:

$$\min (g - H\hat{f})^T(g - H\hat{f}) + \gamma(W\hat{f})^T(W\hat{f}) \quad (17)$$

where the W matrix represents the intensity weighting for the overall smoothness measure of the image, and γ is a proper regularization parameter.

To formulate the regularized image restoration problem in terms of ANN, the key step is to derive an energy function so that the lowest energy state (the most stable state of the network) would correspond to the best restored image. Once the energy function is determined, the synaptic strengths and input can be immediately derived. Let each pixel of image f_i correspond to the neuron state V_i , then the derived energy function is equal to the expression given in Eq. 17. By comparing Eq. 12 and Eq. 17, the corresponding T matrix and i vector are found to be

$$\begin{aligned} T &= -2(H^T H + \gamma W^T W) \\ i &= 2H^T g \end{aligned} \quad (18)$$

Once T and i are determined, the ANN can be programmed accordingly to solve the image restoration problem.

Using ANN for Associative Pattern Recognizer The discrete-state Hopfield model can be considered as an associative memory; therefore, it may be adopted for image pattern recognition applications [39], [Psalt85]. Briefly, the discrete-time transition of neuron i is formulated as (with $I_i = 0$):

$$\begin{aligned} U_i(k) &= \sum_j^N T_{ij} V_j(k) \\ V_i(k+1) &= \text{step}[U_i(k)] \end{aligned} \quad (19)$$

where $\text{step}[x]$ is a unit step function, which is 1 for $x \geq 0$ and 0 for $x < 0$.

For the associative retrieval purpose, let the reference patterns be represented by $v^m = [V_1^m, V_2^m, \dots, V_N^m]$, where $m = 1, 2, \dots, M$, then the synaptic strengths $\{T_{ij}\}$, where $1 \leq i, j \leq N$, are derived as follows:

$$T_{ij} = \begin{cases} \sum_{m=1}^M (2V_i^m - 1)(2V_j^m - 1) & i \neq j \\ 0 & i = j \end{cases} \quad (20)$$

where the states $\{V_i^m\}$ have binary values 0 or 1.

For pattern recognition, an unknown test pattern is initially input to the SFG ANN of Figure 15. The network iterates following the transition rule given in Eq. 19. When the state of the neural network converges to a final state v^* , then v^* is output from the network. In a pattern recognizer, this output v^* must be compared with the M reference patterns $\{v^m, m = 1, 2, \dots, N\}$ to determine if it exactly matches any of the reference patterns. If it does not match then a "no match" is declared. If it does, the output will be the matched reference pattern.

4 Kalman Filtering for Least Squares Estimation

Kalman filtering has found frequent applications to many time varying (or time invariant) signal processing problems, such as those encountered in communications and control, radar and sonar processing and so on. A discrete time-varying recursive dynamic system can be represented as:

$$\begin{aligned} x(k+1) &= F(k)x(k) + w(k+1) \\ y(k) &= C(k)x(k) + v(k) \end{aligned} \quad (21)$$

where $F(k)$ and $C(k)$ are coefficient matrices with dimension $n \times n$ and $m \times n$; $x(k)$ and $w(k+1)$ are the n -dimensional state vector and system noise vector respectively; $y(k)$ and $v(k)$ are the m -dimensional measurement vector and measurement noise vector respectively. The noise vectors have zero mean and known covariance matrices $R_w(k+1)$ and $R_v(k)$ respectively. The noise w is assumed to be uncorrelated with v (i.e. $E[w(i)v(j)] = 0$, for all i, j).

4.1 Least Squares Formulation of Kalman Filtering

The Kalman filtering is to compute the optimal prediction of $x(k+1)$, denoted as $\hat{x}(k+1)$, based on known measurements $\{y(1), y(2), \dots, y(k)\}$ [40]. The best prediction for the state vector in Eq. 21, based on a linear minimum variance criterion, can be solved recursively with computational complexity of $O(n^3)$. In the following, we reformulate the Kalman filtering algorithm into a classic least squares problem. For more details, see [41], [42], [43].

For colored noise cases, where the covariance matrix, $R_w(k+1)$ and $R_v(k)$, are not identity matrices, a whitening procedure has to be first applied. The covariance matrices of the two noise vectors can be expressed as:

$$\begin{aligned} R_w^{-1}(k+1) &= W(k+1)W^T(k+1) \\ R_v^{-1}(k) &= V(k)V^T(k) \end{aligned}$$

where $\mathbf{W}(k+1)$ and $\mathbf{V}(k)$ are *upper triangular matrices* obtained from Cholesky decomposition [32]. Applying premultiplication of the whitening operators $\mathbf{W}(k+1)$ and $\mathbf{V}(k)$ to Eq. 21 yields:

$$\begin{aligned}\mathbf{W}(k+1)\mathbf{x}(k+1) &= \mathbf{W}(k+1)\mathbf{F}(k)\mathbf{x}(k) + \mathbf{W}(k+1)\mathbf{w}(k+1) \\ \mathbf{V}(k)\mathbf{y}(k) &= \mathbf{V}(k)\mathbf{C}(k)\mathbf{x}(k) + \mathbf{V}(k)\mathbf{v}(k)\end{aligned}\quad (22)$$

By grouping together the state vectors up to stage k and forming a large vector $\mathbf{X}(k)$, Eq. 22 can be represented by the following matrix-vector form [41]

$$\tilde{\mathbf{U}}(k) = \tilde{\mathbf{A}}(k)\mathbf{X}(k) + \tilde{\mathbf{Y}}(k) \quad (23)$$

where

$$\begin{aligned}\mathbf{X}(k) &= [\mathbf{x}^T(1) \ \mathbf{x}^T(2) \ \dots \ \mathbf{x}^T(k)]^T \\ \tilde{\mathbf{U}}(k) &= [\tilde{\mathbf{w}}^T(1) \ \tilde{\mathbf{v}}^T(1) \ \tilde{\mathbf{w}}^T(2) \ \tilde{\mathbf{v}}^T(2) \ \dots \ \tilde{\mathbf{v}}^T(k-1) \ \tilde{\mathbf{w}}^T(k)]^T \\ \tilde{\mathbf{Y}}(k) &= [\mathbf{0} \ \tilde{\mathbf{y}}^T(1) \ \mathbf{0} \ \tilde{\mathbf{y}}^T(2) \ \dots \ \tilde{\mathbf{y}}^T(k-1) \ \mathbf{0}]^T\end{aligned}$$

and

$$\tilde{\mathbf{A}}(k) = \begin{bmatrix} \mathbf{W}(1) & & & & & & \\ \tilde{\mathbf{C}}(1) & & & & & & \\ \tilde{\mathbf{F}}(1) & \mathbf{W}(2) & & & & & \mathbf{0} \\ & \tilde{\mathbf{C}}(2) & & & & & \\ & \tilde{\mathbf{F}}(2) & \mathbf{W}(3) & & & & \\ & & & & & & \\ & & & & & \mathbf{W}(k-1) & \\ & & & & & \tilde{\mathbf{C}}(k-1) & \\ & & & & & \tilde{\mathbf{F}}(k-1) & \mathbf{W}(k) \end{bmatrix}$$

where $\tilde{\mathbf{F}}(k) = -\mathbf{W}(k+1)\mathbf{F}(k)$, $\tilde{\mathbf{C}}(k) = \mathbf{V}(k)\mathbf{C}(k)$, $\tilde{\mathbf{w}}(k+1) = \mathbf{W}(k+1)\mathbf{w}(k+1)$, $\tilde{\mathbf{y}}(k) = -\mathbf{V}(k)\mathbf{y}(k)$, and $\tilde{\mathbf{v}}(k) = -\mathbf{V}(k)\mathbf{v}(k)$. In the above, it is assumed that $\mathbf{x}(0) = \mathbf{0}$ and therefore $\tilde{\mathbf{w}}(1) = \mathbf{W}(1)\mathbf{x}(1)$.

The noise vector $\tilde{\mathbf{U}}(k)$ in Eq. 23 is white with zero mean and an identity covariance matrix. Therefore the best estimate, $\hat{\mathbf{x}}(k)$, given $\{\mathbf{y}(1) \dots \mathbf{y}(k-1)\}$, can now be solved as a least squares estimation problem using the QR decomposition method. Applying an orthogonal transformation matrix \mathbf{Q} , with a dimension of $[(k-1)m + kn] \times [(k-1)m + kn]$ at stage k , for both sides of Eq. 23, gives:

$$Q\bar{U}(k) = Q\bar{A}(k)\bar{X}(k) + Q\bar{Y}(k) \quad (24)$$

where

$$[Q\bar{A}(k)|Q\bar{Y}(k)] = \left[\begin{array}{ccccc|c} R_{11} & R_{12} & & & & b'_1 \\ & R_{22} & R_{23} & 0 & & b'_2 \\ & & \ddots & & & \vdots \\ & 0 & & R_{k-1,k-1} & R_{k-1,k} & b'_{k-1} \\ & & & & R(k) & b_k \\ 0 & 0 & 0 & \dots & 0 & r_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & r_{k-1} \end{array} \right]$$

In the above expression the submatrices R_{ii} and $R(k)$ are upper triangular; $\{b'_1, \dots, b'_{k-1}, b_k\}$ are n dimensional vectors; and $\{r_1, r_2, \dots, r_{k-1}\}$ are m dimensional residual vectors. Note also that under the orthogonal transformation, $Q\bar{U}(k)$ remains white and therefore the best estimated state vector $\hat{x}(k)$ depends only on the vector b_k :

$$\hat{x}(k) = -R(k)^{-1}b_k \quad (25)$$

Since $R(k)$ is an upper triangular matrix, Eq. 25 can be solved by back substitution.

4.2 Recursive Least Squares Estimation of Kalman Filter

At the next recursion, with the new measurement $y(k)$, the updated system equation for estimating $\hat{x}(k+1)$, is given by a modified matrix-vector form:

$$\hat{U}(k+1) = \hat{A}(k+1)\hat{X}(k+1) + \hat{Y}(k+1) \quad (26)$$

where

$$\hat{U}(k+1) = \left[\begin{array}{c} Q_1\bar{U}(k) \\ \vdots \\ \tilde{v}(k) \\ \tilde{w}(k+1) \end{array} \right]$$

and

$$[\tilde{\mathbf{A}}(k+1)|\dot{\mathbf{Y}}(k+1)] =$$

$$\left[\begin{array}{cccccc|c} \mathbf{R}_{11} & \mathbf{R}_{12} & & & & & \mathbf{b}'_1 \\ & \mathbf{R}_{22} & \mathbf{R}_{23} & & & & \mathbf{b}'_2 \\ & & \ddots & & & & \vdots \\ & 0 & & \mathbf{R}_{k-1,k-1} & \mathbf{R}_{k-1,k} & & \mathbf{b}'_{k-1} \\ & & & & \mathbf{R}(k) & & \mathbf{b}_k \\ 0 & 0 & \dots & 0 & 0 & 0 & \mathbf{r}_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \mathbf{r}_{k-1} \\ \hline 0 & 0 & \dots & 0 & \tilde{\mathbf{C}}(k) & 0 & \tilde{\mathbf{y}}(k) \\ 0 & 0 & \dots & 0 & \tilde{\mathbf{F}}(k) & \mathbf{W}(k+1) & 0 \end{array} \right]$$

To compute $\hat{\mathbf{x}}(k+1)$ by QR decomposition on $\tilde{\mathbf{A}}(k+1)$ we need only be concerned with the following $(2n+m) \times (2n+1)$ matrix,

$$\left[\begin{array}{ccc|c} \mathbf{R}(k) & 0 & | & \mathbf{b}_k \\ \tilde{\mathbf{C}}(k) & 0 & | & \tilde{\mathbf{y}}(k) \\ \tilde{\mathbf{F}}(k) & \mathbf{W}(k+1) & | & 0 \end{array} \right] \quad (27)$$

as opposed to the large $\tilde{\mathbf{A}}(k+1)$ matrix. Performing the QR decomposition on Eq. 27 using the orthogonal transformation matrix leads to the following:

$$\mathbf{Q}_1 \left[\begin{array}{ccc|c} \mathbf{R}(k) & 0 & | & \mathbf{b} \\ \tilde{\mathbf{C}}(k) & 0 & | & \tilde{\mathbf{y}}(k) \\ \tilde{\mathbf{F}}(k) & \mathbf{W}(k+1) & | & 0 \end{array} \right] = \left[\begin{array}{ccc|c} \mathbf{R}_{k,k} & \mathbf{R}_{k,k+1} & | & \mathbf{b}'_k \\ 0 & \mathbf{R}(k+1) & | & \mathbf{b}'_{k+1} \\ 0 & 0 & | & \mathbf{r}_k \end{array} \right] \quad (28)$$

where $\mathbf{R}_{k,k}$ and $\mathbf{R}(k+1)$ are $n \times n$ upper triangular matrices, and \mathbf{r}_k is a residual vector.

SFG Array Design of Kalman Filter Array processor designs for the Kalman filter have been proposed in [44], [45], [46], [43]. The results in [44] and [45] handle only the measurement updating, while [46] and [43] propose array designs for both the measurement and time updatings. However, all the previous methods fail to exploit the inherent triangular structure of the matrix $\mathbf{W}(k+1)$ in Eq. 27. Now we take advantage of this special structure to derive a more efficient SFG design.

Algorithm Description The overall triangualization procedure consists of two parts

1. Nullification of $[\tilde{\mathbf{C}}^T(k)\tilde{\mathbf{F}}^T(k)]^T$ by rotation with $\mathbf{R}(k)$ in a QR triarray (see Example 2).
2. Update $\mathbf{W}(k+1)$ while retaining its triangular structure.

On the surface, the first step appears to be straightforward. However, as shown in Figure 16, any rotation operations applied to each pair of rows in the matrices $\tilde{F}(k)$ and $R(k)$ are also applied to the corresponding pair of rows in the matrices $W(k+1)$ and the zero matrix above it, denoted by $Z(k)$. In order to guarantee a successful preservation of the upper triangular structure of $W(k+1)$ in the second step, some processing ordering must be adopted when performing operations on $\tilde{F}(k)$ in the first step. This can be illustrated by showing how the substeps progress and displaying their snapshots in Figure 16. We note that the $\tilde{F}(k)$ matrix is nullified in a bottom up order, as shown in Figure 16, hence allowing $W(k+1)$ to remain upper triangular throughout entire process.

SFG Architecture Design The above two-step procedure can be performed by a single QR triarray. The operation is divided into two phases: the processing on the $\tilde{C}(k)$ and $\tilde{F}(k)$ matrices is performed in the first phase (see Figure 17(a)); while the processing on $W(k+1)$ is performed in the second phase (see Figure 17(b)). The processing time for each recursion (i.e., both phases) is $4n + m$ [13]. The following describes how the triarray works:

1. **Operation on $\tilde{C}(k)$:** The new arriving matrix $\tilde{C}(k)$ can be nullified by rotating with the resident triangular matrix $R(k)$ (see Figure 17(a)).

Operation on $\tilde{F}(k)$: As shown in Figure 17(a), the nullification on $\tilde{F}(k)$ will continue right after (m time units) the operation on $\tilde{C}(k)$ in a similar manner in the triarray. As discussed in Example 2, a diagonal PE performs GG (Givens generation) and then sends the GR (Givens rotation) parameters right-ward to the remaining PEs in the same row, where the GRs are performed. Most importantly, after $n + m$ time units, the parameters of the rotation angles $\{\theta_{ij}\}$, (i.e., $\{\cos\theta_{ij}, \sin\theta_{ij}\}$), start to emerge from the right side of the triarray and are stored in a data buffer to be used for the purpose of the second phase processing.

2. **Operation on $W(k+1)$:** We will prove in a moment that, right after the nullification of $\tilde{F}(k)$ is completed at $t = 3n + m$, the matrix $W(k+1)$ is already loaded into the triarray and the rotation parameters of $\{\theta_{ij}\}$ also become available for use. Given what we have just claimed, it is straightforward to perform the rotation operation on $W(k+1)$ in a triarray. Note that the triarray should now provide upward and rightward communication channels, as opposed to the downward and rightward ones provided in the first phase processing.

Rotation Parameters: As shown in Figure 18(a), the parameters of the rotation angles, $\{\theta_{ij}\}$, emerge from the right side of the triarray in a skewed data pattern. The data along any anti-diagonal line will be sent to the same diagonal PE in the triarray. A natural design is to use a data buffer to store these parameters, as shown in Figure 18(b), with the top row of the data buffer directly linked to the diagonal PEs of the triarray. At $t = 3n + m$, with all the parameters of $\{\theta_{ij}\}$ in place, they are sent row by row to the triarray to perform the Givens rotations on $W(k+1)$.

$$\begin{array}{c}
 \left[\begin{array}{c|c} \frac{x(k)}{y(k)} & \vdots \\ \hline \frac{z(k)}{w(k+1)} & \end{array} \right] = \left[\begin{array}{c|c} \frac{x^*(k)}{y(k)} & \frac{z(k)}{w(k+1)} \\ \hline \frac{z(k)}{w(k+1)} & \end{array} \right] = \\
 \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \quad \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \\
 \downarrow \qquad \qquad \downarrow \\
 \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \quad \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \\
 \downarrow \qquad \qquad \downarrow \\
 \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \quad \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \\
 \downarrow \qquad \qquad \downarrow \\
 \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \quad \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \\
 \downarrow \qquad \qquad \downarrow \\
 \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right] \quad \left[\begin{array}{c|c} \vdots & \vdots \\ \hline \vdots & \vdots \end{array} \right]
 \end{array}$$

Figure 16: The progress of the nullification of the $W(k+1)$ matrix, while retaining its triangular structure, where the blank elements are also regarded as 0. The \circ elements indicate the active row elements involved in that substep of the Givens rotation, and a 0 indicates where nullification is taking place.

The overall data arrangements for the triarray are displayed in Figure 19. The time lag for the $\{\theta_{ij}\}$ parameters reflects the time for the triarray and buffer processing.

Loading $W(k+1)$: As shown in Figure 19, $W(k+1)$ is loaded (from the top) into the triarray immediately after $\tilde{F}(k)$. This allows perfect synchrony in a timely engagement with the $\{\theta_{ij}\}$ parameters (from the left) and $Z(k)$ data (from the bottom), (see Figure 19).

Solution of $\hat{x}(k+1)$ The best prediction of the state vector $\hat{x}(k+1)$ is obtained by solving Eq. 25. This is easily done using the back substitution method on one additional linear array of length n .

Processing Time for One Recursion There are two phases in the operation. The first phase takes $3n + m$ time units and the second phase takes $2n$ time units to complete their respective operations. However, immediately after n time units of the second phase processing, PE(1,1) becomes free and can be used to start processing the new recursion (for the $(k+1)$ st stage). Thus the processing time for each recursion is only $4n + m$ (instead of $5n + m$).

More elaborately, as shown in Figure 19, the k -th stage processing starts at $t = 0$ when $C_{m1}(k)$ enters the array, while the $(k+1)$ st stage processing (for $C_{m1}(k+1)$) can start right after Z_{41} finishes its rotation at PE(1, 1) at $t = 4n + m$. (Note that at this time PE(1, 2) will still be performing the stage k computation but it will become available for the stage $k+1$ at the next time unit when C_{m2} enters the array.) This amounts to $4n + m$ time units for processing one recursion [13].

5 Concluding Remarks

The signal and image processing applications generally call for algorithms which are deterministic in both time and space. This allows development of an unified theoretical framework for architecture and algorithm optimization for optical/VLSI array processors. Indeed, *algorithmically specialized parallel computers* will very soon become a dominant and promising trend in future supercomputing technology. We will witness the construction of an increasing number of specialized array processors/computers tailored to execute important *kernel* algorithms from signal/image processing, numerical linear algebra, and scientific computing. Therefore, parallel algorithm analysis, design, and mapping methodology onto optical/VLSI array processors will have a vital role. (Note that the SFG arrays proposed here can be further mapped to systolic arrays using the mapping methodology proposed by S.Y. Kung [18].)

In the near future, we will face the challenge of modern and *intelligent* supercomputer systems, which are required to perform perception-level functions. These systems will be more adaptive to their environment and interact with users in more natural and efficient ways through the integration of logic circuitry and sensor stimulators. By launching novel

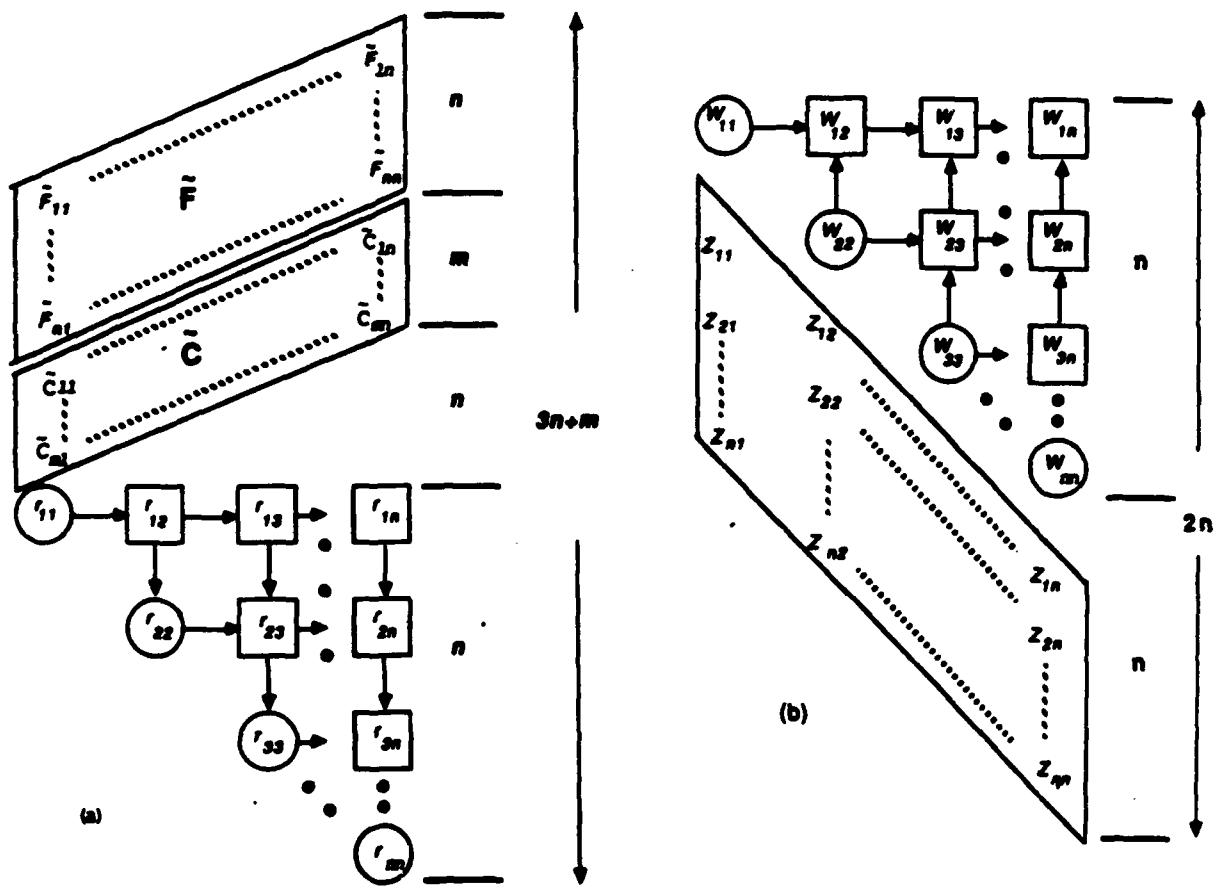


Figure 17: The two steps of one recursion of Kalman filter updating. (a) Using $R(k)$ to nullify the $\tilde{C}(k)$ and $\tilde{F}(k)$ matrices. (b) Rotating $W(k+1)$ and $Z(k)$.

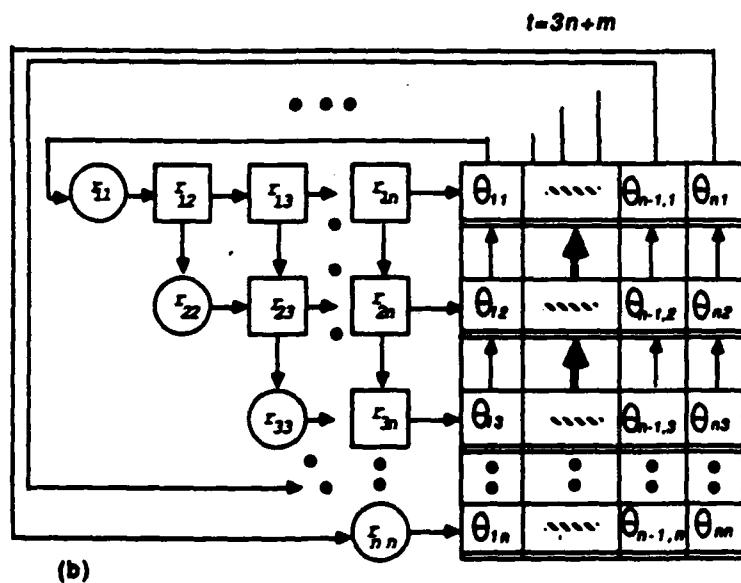
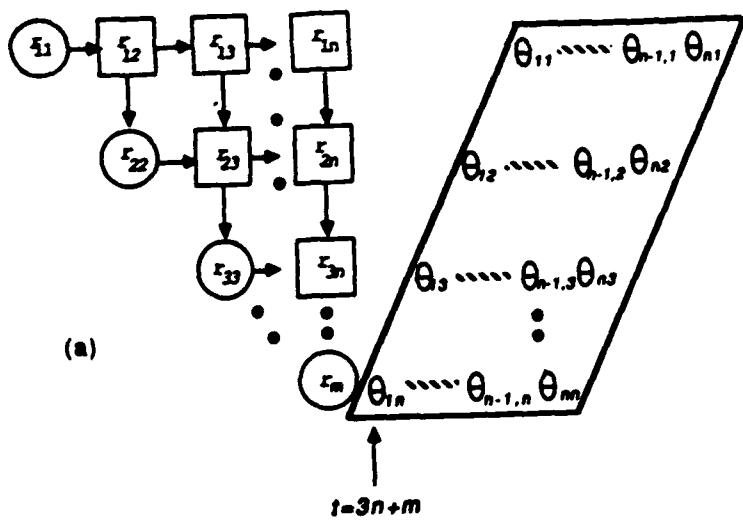


Figure 18: (a) The parameters of the rotation angles, $\{\theta_{ij}\}$, emerge from the right side of the triarray in a skewed data pattern. (b) Data buffer to store these parameters, with the top row directly linked to the diagonal PEs of the triarray.

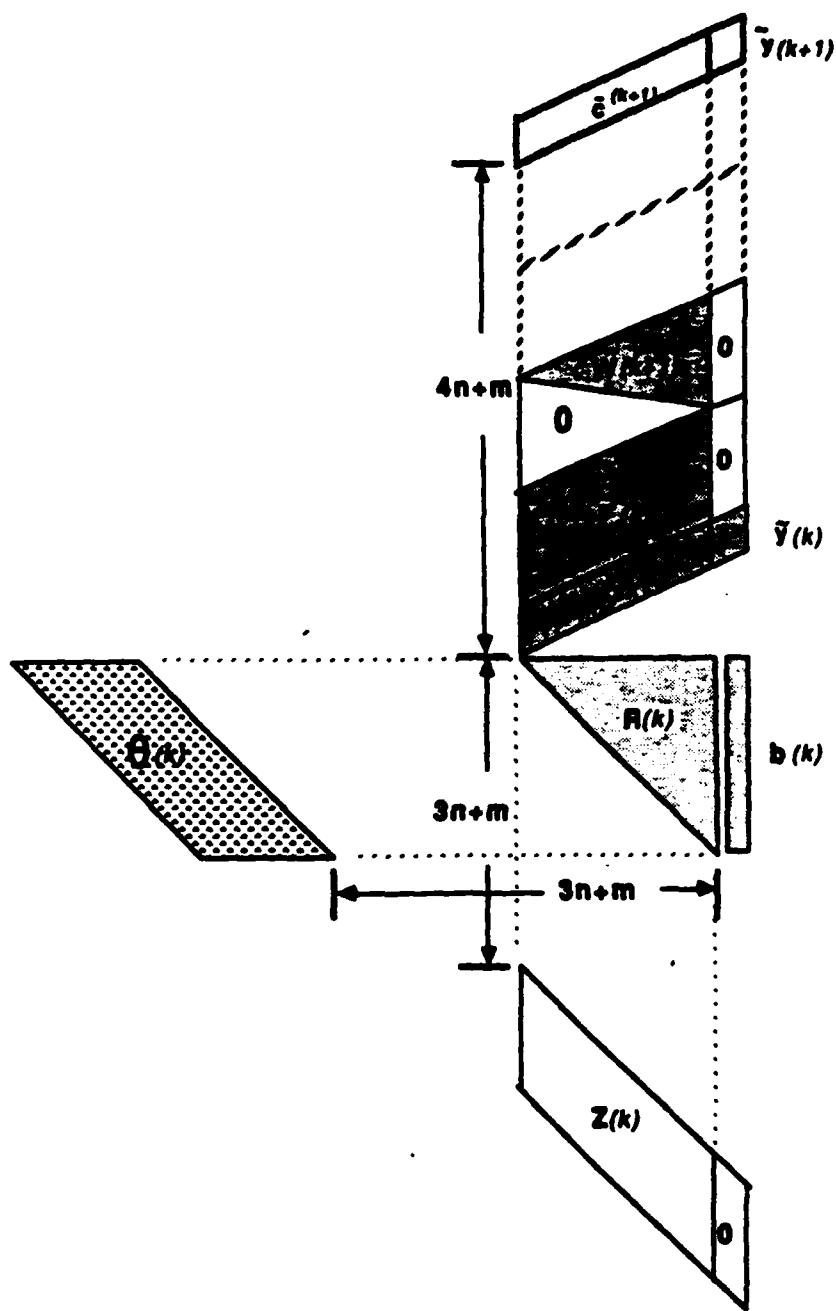


Figure 19: The overall data arrangements for one recursive updating in the Kalman filter algorithm.

computing concepts based on optical/VLSI technologies, these systems are also supposed to solve otherwise unsolvable problems. From the perspective of such systems, the massively parallel array processors proposed in this paper can only be regarded as basic primitives, upon which novel computing (or thinking) machines are yet to be built.

References

- [1] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statist. Soc., series B*, 48(3):192-236, 1986.
- [2] J. Besag. Spatial interaction and statistical analysis of lattice system (with discussion). *J. Royal Statist. Soc., series B*, 36:192-236, 1974.
- [3] J.-M Delosme and I. C. F. Ipsen. Efficient systolic arrays for the solution of Toeplitz systems: an illustration of a methodology for the construction of systolic architectures in VLSI. *International Workshop on Systolic Arrays, University of Oxford*, F2, July 1986.
- [4] A. L. Fisher. Systolic algorithms for running order statistics in signal and image processing. *J. of Digital System*, 4:251-264, 1982.
- [5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6:721-741, November 1984.
- [6] J.P. Hayes. *Computer Architecture and Organization*. McGraw Hill, N.Y., 1978.
- [7] R. M. Karp, R. E. Miller, and S. Winograd. The organization of computations for uniform recurrence equations. *Journal of ACM*, 14(3):563-590, July 1967.
- [8] S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi. Optimization by simulated annealing. *Science*, May 1983.
- [9] S. Y. Kung, P. S. Lewis, and S. C. Lo. On optimally mapping algorithms to systolic arrays with application to the transitive closure problem. In *Proc. 1986 IEEE Int. Sym. on Circuits and Systems*, pp 1316-1322, 1986.
- [10] S. Y. Kung, P. S. Lewis, and S. N. Jean. Canonic and generalized mapping from algorithms to arrays - a graph based methodology. In *Proc. of the Hawaii Inter. Conf. on System Sciences*, Vol. 1, pp. 124 - 133, January, 1987.
- [11] S. Y. Kung and J. N. Hwang. Systolic design of electronic neural networks. Submitted to *26th IEEE Conf. on Decision and Control*, Los Angeles, December, 1987.
- [12] S. Y. Kung, J. N. Hwang, and S. C. Lo. Systolic/Wavefront arrays for image processing algorithms. *Mari87, Intelligent Networks and Machines*, France, May, 1987.
- [13] S. Y. Kung, J. N. Hwang, S. C. Lo, R. W. Stewart. A new systolic array for Kalman Filtering. Submitted to *19th Annual Asilomar Conf. on Signals, Systems and Computers*, November 1987.
- [14] S. Y. Kung, R. W. Stewart, S. C. Lo. Systolic Array Processing Image/Video Processing Applications *SPIE Visual Communications and Image Processing II, Cambridge Massachusetts, October 1987*

- [15] J. Vlontzos, and S. Y. Kung. A wavefront array processor using dataflow processing elements. In *Proc. Int'l' Conf. on Supercomputing*, Athens, Greece, June, 1987.
- [16] H.T. Kung. Why systolic architectures? *IEEE, Computer*, 15(1), Jan 1982.
- [17] H.T. Kung and C.E. Leiserson. Systolic arrays (for VLSI). In *Sparse Matrix Symposium*, pages 256–282, SIAM, 1978.
- [18] S. Y. Kung. *VLSI Array Processors*. Prentice Hall, 1987.
- [19] S. Y. Kung, J. N. Hwang, and S. C. Lo. Mapping digital signal processing algorithms onto VLSI systolic/wavefront arrays. In *Proc. 12th Annual Asilomar Conf. on Signals, Systems and Computers*, November 1986.
- [20] S. Y. Kung, S. C. Lo, and J. Annevelink. Temporal localization and systolization of signal flow graph (SFG) computing networks. In , SPIE, McGregor and Werner, Inc., San Diego, CA, August 1984.
- [21] S.Y. Kung. On supercomputing with systolic/wavefront array processors. *Invited paper, Proceedings of the IEEE*, Vol. 72(7), July 1984.
- [22] S.Y. Kung, K.S. Arun, R.J. Gal-Ezer, and D.V. Bhaskar Rao. Wavefront array processor: language, architecture, and applications. *IEEE Transactions on Computers, Special Issue on Parallel and Distributed Computers*, C-31(11):1054–1066, Nov 1982.
- [23] C. C. Kuo. Parallel algorithms and architectures for solving elliptic partial differential equations. MIT Technical Report (LIDS-TH-1432), January, 1985.
- [24] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of Royal Society of London*, 207:187–217, 1980.
- [25] K. Oflazer. Design and implementation of a single chip 1-D median filter. *IEEE Transactions on ASSP*, October 1983.
- [26] I.N. Parker. VLSI architecture. Book: *VLSI Image Processing*, 99–127, 1985.
- [27] W. K. Pratt. *Digital Image Processing*. Wiley, 1978.
- [28] I. Offen and R.J. Raymond Jr. *VLSI Image Processing*. McGraw-Hill, 1985
- [29] J. J. Hopfield. Neural Network and Physical Systems with Emergent Collective Computational Abilities. *Proc. Natl. Acad. Sci. USA*, 1982
- [30] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, Vol. 81, Pages 3088–3092, 1984
- [31] J. J. Hopfield and D. W. Tank. Neural Computation of Decision in optimization Problems. *Biological Cybernetics*, 1985
- [32] G.H. Golub and C. van Loan. *Matrix Computations* Johns Hopkins University Press, 1983.
- [33] M. Takeda and J.W. Goodman. Neural networks for computation: number representations and programming complexity. *Applied Optics*, Vol. 25: pp. 3033–3046, September 1986.
- [34] N. H. Farhat, D. Psaltis, A. Prata and E. Paek. Optical implementation of the Hopfield model. *Applied Optics*, Vol. 24: pp. 1469– 1475, May 1985.

- [35] H. Mada. Architecture for optical computing using holographic associative memories. *Applied Optics*, Vol. 24: 1985.
- [36] G. Eichmann and H. J. Caulfield. Optical learning (inference) machines. *Applied Optics*, Vol. 24: 1985.
- [37] C. Koch, J. Marroquin and A. Yuille. Analog "neuronal" networks in early vision. *Proc. of National Academy Science*, Vol. 83: pp. 4263-4267, 1986.
- [38] W. A. Porter, and J. L. Aravena. Array architectures for estimation and control applications: An introduction and overview. Submitted to *25th IEEE Conf. on Decision and Control*, Athens, 1986.
- [39] L. E. Atlas, T. Homma, R. J. Marks II. A neural network model for vowel classification. Presented in *Proc. IEEE ICASSP'87*, Dallas, Texas, 27.15.1-27.15.4, 1987.
- [Psalt85] D. Psaltis and N. Farhat. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Optics Letters*, 10: pp. 98-99, Jan., 1985.
- [40] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Engineering*, 82: pp 35-45, 1960.
- [41] C. C. Paige and M. A. Saunders. Least squares estimation of discrete linear dynamic systems using orthogonal transformation. *SIAM J. Numer. Anal.*, 14: pp. 180-193, 1977.
- [42] M. J. Chen and K. Yao. On realization of least-squares estimation and Kalman filtering by systolic arrays. In *Proc. of Intl. Workshop on Systolic Arrays*, Oxford, July 1986.
- [43] M. J. Chen and K. Yao. On realization and implementation of Kalman filtering by systolic array. In *Proc. of John Hopkins Workshop*, 1987.
- [44] A. Andrew. Parallel processing of the Kalman filter. In *Proc. Int. Conf. Parallel Processing*, pp. 216-220, 1981.
- [45] J. M. Jover and T. Kailath. A parallel architecture for Kalman filter measurement update and parameter estimation. *Automatica*, 22: pp. 43-57, 1986.
- [46] T. Y. Sung and Y. H. Hu. VLSI implementation of real-time Kalman filter. In *Proc. IEEE ICASSP'86*, Tokyo, Japan, pp. 2223-2226, 1986.